A METHODOLOGY FOR SEPARATION OF MULTIPLE DISTRIBUTIONS IN ARTERIAL TRAVEL TIME DATA

A Thesis
Presented to
The Academic Faculty

by

James Miller Anderson, IV

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Civil Engineering in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
August 2015

A METHODOLOGY FOR SEPARATION OF MULTIPLE DISTRIBUTIONS IN ARTERIAL TRAVEL TIME DATA

Approved by:

Dr. Michael Hunter, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Michael Rodgers
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Angshuman Guin
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: May 15, 2015

ACKNOWLEDGEMENTS

I owe a debt of gratitude to many people who have dedicated their time and patience in the development of this thesis. This thesis would not be possible without the dedication, support, and guidance of many Georgia Tech Professors, Research Engineers, fellow, and former students.

Many thanks go to the undergraduate researchers and graduate researchers who have been generous with their time in assisting with data collection efforts for this project. Especially, Prabha Pratyaksa for his generous help in all things research and non-research. I would also like to thank Sarah Windmiller for her unwavering moral support through my last year of graduate school.

I especially would like to thank Dr. Wonho Suh for his support and helpfulness in teaching me the ins and outs of microscopic traffic simulation, VBA scripting, and also for all the lunches of the last several years. Special thanks also goes to Dr. Angshuman Guin for his time and efforts in providing his expertise with technology to consistently solve problems, access data sources, and aid in gathering data crucial to this project.

Finally, I would like to thank Dr. Michael Hunter and Dr. Michael Rodgers for their dedication and patience in development of this thesis. Dr. Hunter especially for giving excellent guidance with coordinated signal timing plans, other signal timing oddities, microscopic simulation, and simulation calibration. And, Dr. Rodgers especially for his patience in answering all of my questions regarding statistics as well as having the patience to help me understand the specifics of the statistical analyses undertaken in this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

EM	Expectation Maximization
R	R-Statistical Language
CRAN	Comprehensive R Archive Network
DDI	Diverging Diamond Interchange
GDOT	Georgia Department of Transportation
MAC	Media Access Control
GHz	Gigahertz
NB	Northbound
SB	Southbound
ALPR	Automated License Plate Recognition
AIC	Akaike Information Criterion
MPH	Miles Per Hour
FIFO	First In First Out
LILO	Last In Last Out

SUMMARY

Multiple distribution travel time data has been observed in signalized corridors as well as freeway corridors. This behavior is typically caused by congestion, uncoordinated signals, or routes through a coordinated corridor that are not a priority. On the SR140 corridor near the Jimmy Carter Boulevard / I-85 Interchange, it was found that the travel times recorded on the corridor contained multiple distributions and thus a methodology was sought to properly separate the distributions in order to perform more robust statistical analysis.

Next, an R statistical language library was found, called “mixtools”, which contained a multiple gamma distribution fitting function called “gammamixEM”. Gamma distributions were chosen for this application as typical travel time distributions tend contain a one sided tail. This function was used in conjunction with a monte-carlo approach to find fits for one to six distributions. The accuracy of the fit was confirmed through visual inspection of the plotted distributions. Then, the Akaike Information Criteria were used to compare the fits to determine the best fit number of distributions.

This thesis contains a detailed outline of the algorithm as well as results from the algorithm for the combined Tuesday dataset from this project. It was found that the approach worked well for 60 out of 70 cases. In the 10 cases that were not ideal, the distributional fits make sense on a statistical level, however, for the purposes of the before and after project the next best Akaike Information Criteria value fit may need to used. These 10 cases tended to split obvious single distributions into two distributions, which is not desirable in a before and after analysis where one is not only testing individual distributions before and after construction but also determining if distributions were created or removed as a result of the change in operation of the interchange.

CHAPTER 1

INTRODUCTION

1.1 Previous Research

Multi-distributional travel times have been observed in a number of settings. These include arterials [1, 2], due to the presence of signalized intersections, and freeways [3-5] when there are lane closures during construction periods or, more generally, when one is observing travel times over a number of hours containing both congested and uncongested travel. These multi-distributional travel times reflect different subgroups of traffic which may experience different service quality along the same roadway. Previous studies have used Expectation Maximization and hierarchical Bayesian mixture models to separate these distributions in order to develop travel time reliability indices [1-4]. These methods are becoming more useful in the transportation field as the availability of disaggregate data continues to grow from advancing technologies such as Bluetooth® travel time detectors, Automated License Plate Recognition (ALPR) cameras, cell phone tracking, toll tag readers, and magnetic signature detectors.

1.2 Data Source, Collection Sites, and Routes

The data used in this research was collected using commercial "off–the-shelf" Bluetooth® travel time detection equipment. A total of six Bluetooth® detectors were deployed in the field by the Georgia Department of Transportation (GDOT) over the study area. These detectors recorded Bluetooth® Media Access Control (MAC) addresses of Bluetooth® equipped devices (in discoverable mode) within vehicles traversing the study location. MAC addresses are assigned under a scheme designed to reduce the likelihood that any two devices will have the same MAC address. Under this scheme, 48-bit MAC addresses are comprised of six sets of two alpha numeric pairs. The first three pairs are assigned to a specific hardware manufacturer while the last three are generated by the manufacturer to be unique among all devices they manufacture [6]. Since MAC addresses should be unique among all digital devices, the off the shelf Bluetooth® detectors can upload these MAC addresses to a secure server via a cellular connection and match the MAC addresses between different locations. For the before and after DDI analysis, 20 unique routes were defined for MAC address matching, however, only 10 of the routes were used for this multiple distribution analysis. The 10 routes chosen for analysis overlap the other 10 routes that were defined and cover a longer route. This serves two purposes. First, these 10 routes cover the entire coordinated signal corridor allowing evaluation of the corridor as a whole. Second, the origins and destinations on these routes are farther from each other reducing the impact of error from inexact detection location of the Bluetooth® system. Figure 1 below shows the location of Jimmy Carter Boulevard with respect to Atlanta, GA. Figure 2 shows five maps outlining the 10 routes through the interchange for which Bluetooth® travel times were collected for this analysis.

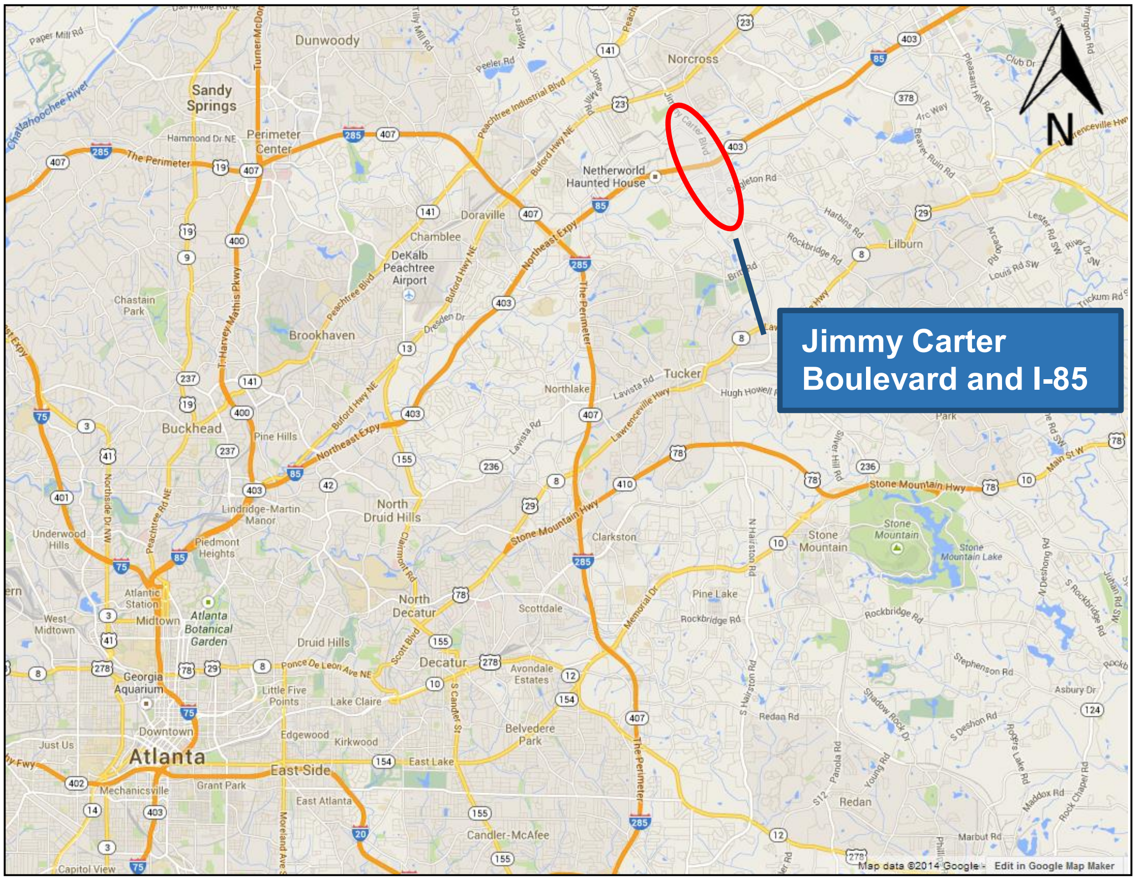


Figure 1: Location of Jimmy Carter Boulevard with respect to Atlanta, GA

Figure 2: The 10 travel time routes used for multiple distribution analysis, (a) Northbound (NB) and Southbound (SB) routes through the interchange (b) NB and SB right turn onto I-85, (c) NB and SB left turn onto I-85, (d) right turns originating from I-85, and (e) left turns originating from the interchange

The 10 routes shown above in Figure 2 are the 10 possible paths that can be taken through the interchange. The routes are shown in green or red lines/arrows and Bluetooth® detectors are shown as blue circles. Figure 2a are the Northbound (NB) and Southbound (SB) routes through the interchange that originate and terminate at the northernmost and southernmost Bluetooth® detectors. Figure 2b shows the two routes originating at the northernmost and southernmost Bluetooth® detectors, making a right turn onto I-85, and terminating at the Bluetooth® detectors located at the interchange ramps. Figure 2c shows the two routes originating at the northernmost and southernmost Bluetooth® detectors, making a left turn onto I-85, and terminating at the Bluetooth® detectors located at the interchange ramps. Figure 2d shows the two routes originating from the I-85 exit ramps, turning right onto Jimmy Carter Boulevard, and terminating at the northernmost and southernmost Bluetooth® detectors. Finally, Figure 2e shows the two routes originating from the I-85 exit ramps, turning left onto Jimmy Carter Boulevard, and terminating at the northernmost and southernmost Bluetooth® detectors.

It can be seen in Figure 2 that only one detector is being used at each origin and destination to detect both directions of travel. At each of these locations the detector is mounted roadside on various objects such as utility poles, signal support poles, or road sign gantry supports. At the interchange ramp locations, the detectors are detecting vehicles not only at the adjacent ramp but also at the ramp located across I-85 from the detector. This large distance (approximately 220ft at the NB on ramp and 250ft at the SB on ramp) is within the range claimed by the commercial Bluetooth® detector manufacturer of 1000ft.. However, due to the larger distances, it is expected that the detection rate of routes originating at the I-85 exit ramps will be lower than that of the other routes originating and terminating at locations closer to the detectors. Besides distance from the detector, other factors can affect the detection rate of Bluetooth® devices.

1.3 Bluetooth® Characteristics Affecting Travel Times

Bluetooth® is a method of wireless communications operating in the 2.4 Gigahertz (GHz) to 2.483 GHz frequencies. The connection process between two Bluetooth® devices can take up to 10.24 seconds. Some studies have found that most devices take between 5.12 and 5.76 seconds to establish a connection [7, 8]. Due to the variance in delay of a connection between devices and the detector and device signal strength and sensitivity different devices may be detected at different distances from the detector. Furthermore, once a device is detected it is typically re-detected several times before the device leaves the detection zone. The detectors used in this study are proprietary and specifications of the detectors as well as their filtering processes for multiple detections are not readily available. Therefore, it is not readily known the exact location where each Bluetooth® device was detected which adds some slight variability to the travel times used in this study. For data verification purposes, license plate data were recorded along this corridor via Automated License Plate Recognition (ALPR) cameras at locations approximately matching the locations of the Bluetooth® sensors' detection areas. The license plates were matched using a Visual Basic script and travel times derived by subtracting the time stamps of the license plate records at each location. The travel times from the ALPR system and the Bluetooth® system were then compared visually. In Figure 3, one can see the Bluetooth® travel times as well as the ALPR travel times plotted against each other. The blue triangles are Bluetooth® data that has been filtered via the commercial Bluetooth® systems filters, the green plus signs are the Bluetooth® raw matches, and the red circles are the ALPR travel times. One can see from Figure 3 that although ALPR generates more travel time data, the commercial Bluetooth® closely matches the ALPR in terms of both the filtered and unfiltered data. Furthermore, both systems show similarities in outlying data points. Finally, this comparison also shows that with enough data points collected on a single day the multiple distribution tendency is prominent. In the ALPR data shown here, one can see two prominent bands from 2:00PM to approximately 4:15PM, with the lower band beginning to diminish, but not completely disappear, as congestion sets in after 4:15PM.

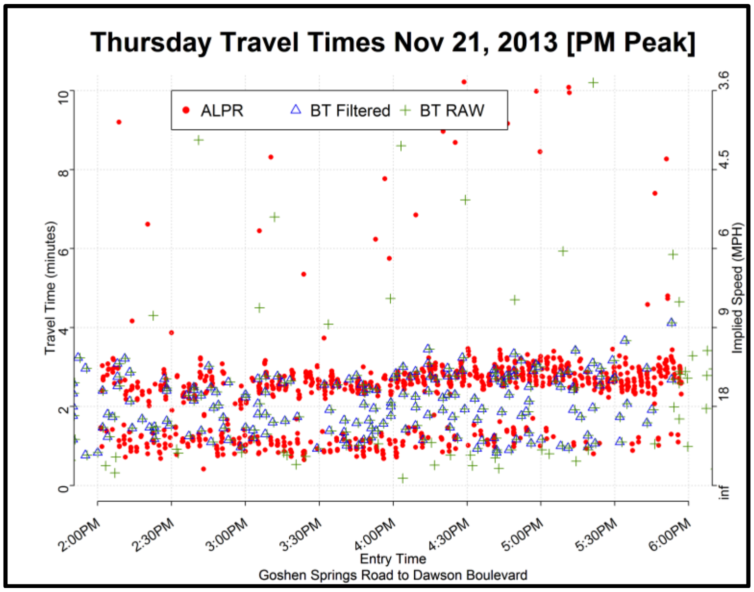


Figure 3: Bluetooth® raw matches, Bluetooth® filtered matches, and ALPR travel time comparison

Besides the variability in detection location effecting travel times, previous studies of Bluetooth® travel time technology have found detection rates ranging from 5–10 percent [9, 10, 11, 12]. These detection rates are calculated as the number of unique MAC address detections divided by the total volume passing the detector. Due to this low detection rate, data were combined across 13 days ranging from August 5th, 2013 to November 22nd, 2013. This is a range of 16 weeks however two weeks were removed from the data due to holidays and one week was removed due to missing data and travel time data that visually appeared not show the same pattern as the remaining 13 weeks of data. Data were not combined across days (e.g. combining data from Monday, Tuesday, Wednesday etc.), however, for each route five data sets were created for each weekday, each day consisting of 13 separate days of data (e.g. a Tuesday dataset consists of 13 different Tuesdays). The individual data points of the mixed data were plotted with each unique day within the data set using a different color symbol so as to ensure that the final data set was well mixed showing that each of the days experienced similar traffic conditions. A sample plot can be seen in Figure 4 below.

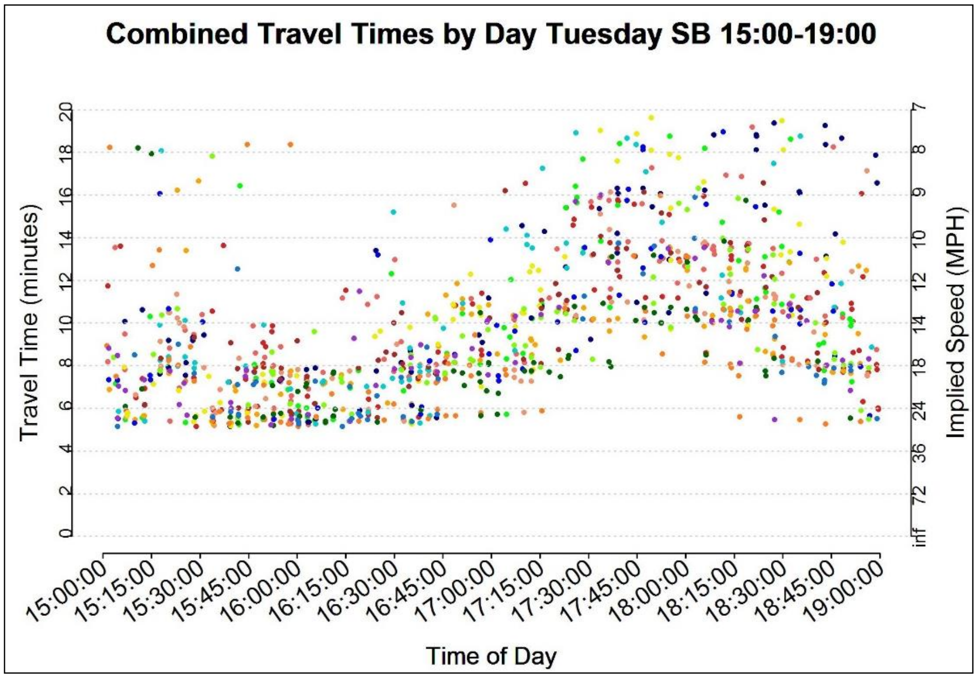


Figure 4: Mixed Tuesday raw data plot from 3PM to 7PM for the SB through route

1.4 Observations in Current Before and After Study

Using data collected from Jimmy Carter Boulevard in Northeast metropolitan Atlanta it was found that travel times along the corridor contained multiple distributions. This behavior is more prevalent during congested peak periods, however, is still found, to a lesser extent, to exist in uncongested conditions. During congested peak periods this behavior is believed to be caused by either vehicle queuing through multiple intersections or bottlenecks and lane speed differentials created by turn bay queue overflows. Under these conditions, as queues grow from the first intersection to the second intersection it becomes more likely that vehicles will be delayed at the second intersection as well as the first intersection due to cycle failure. As queuing continues to grow through multiple intersections travel times begin to jump into higher bands with each intersection captured in the queue. Furthermore, bottlenecks and lane speed differentials created by turn bay queue overflows can cause vehicles on some routes to be caught in the queue for a short period of time, or to reduce their desired speed because of the lane speed differential. The multiple distribution behavior can be seen during uncongested conditions, It is believed there may be two causes of this behavior, either slower moving vehicles operating outside of the green band in the time space diagram of the signal progression or vehicles making short stops along the corridor.

In order to conduct an accurate case study analysis it was decided that fitting a single distribution to the travel time data for statistical analysis would not prove accurate and a multiple distribution method would need to be used. By fitting multiple distributions one can not only observe and record any shifts in each underlying distribution but also detect if any distributions were added or removed. Furthermore, one can individually test each of these distributions to report if the changes in travel times for the distributions were significant. In the case of Jimmy Carter Boulevard this would test whether changes in travel times in each distribution were significant and thus if the change from a traditional diamond interchange to a DDI was effective in reducing travel times and helping mitigate congestion.

CHAPTER 2

METHODOLOGY

2.1 GammamixEM function in R Statistical Programming Language

Development of our own Expectation Maximization (EM) algorithm for fitting multiple distributions is not part of this thesis. These algorithms are a focus of mathematics in their own right and it was our desire to apply an existing tool to a unique problem commonly encountered in the traffic engineering field. Therefore, an existing EM algorithm for fitting multiple distributions was found in the R statistical programming language (herein referred to as R) Comprehensive R Archive Network (CRAN). This algorithm is part of the "mixtools" library in R statistical programming language. This library is not part of the base R install and was developed separately by Benaglia et. al. under a National Science Foundation Grant. The mixtools library contains a number of individual functions, many of which use EM algorithms to fit different types of distributions [14]. For this research, the "gammamixEM" function was used to fit gamma distributions to the data sets [14].

2.1.1 Inputs to the gammamixEM function

Upon investigating the gammamixEM function it was found that there were nine possible inputs to the function two of which are required and the rest set to defaults or chosen at random. These inputs include *x, lambda*, *alpha*, *beta*, *k*, *epsilon*, *maxit*, *maxrestarts*, and *verb*. *X* and *k* are required inputs; *x* being the vector of data points to fit the distribution to and *k* being the number of distributions to fit. The other inputs are not required, however, giving some inputs or changing some of the default values can help increase the stability of the algorithm. *Lambda* is an optional input that is a vector of length equal to *k* that provides the function with a starting point for determining the final proportions of the data between the distributions. *Alpha* is also an optional input vector equal to the length of *k* that provides starting values for the gamma equation variable of alpha, also known as the shape parameter. *Beta* is also an optional vector equal to the length of *k* that provides starting values for the gamma equation variable of beta, also known as the scale parameter. *Epsilon* is the convergence criteria of which the change in log-likelihood from one iteration to the next must be less than for the algorithm to terminate by convergence, its default value is 1*10^-8. *Maxit* is the maximum number of iterations that the algorithm is allowed to try before terminating due to non-convergence, its default value is 1000 iterations. *Maxrestarts* is the maximum number of times the algorithm may restart the iteration counter and choose new starting values if it detects that the initial starting values is not converging quickly enough, its default is set to 20. Finally, *verb* is a true or false input turning on or off the verbal arguments displaying the current iterations log-likelihood, iteration count, and log-likelihood difference, this input's default is false giving no verbal arguments [14].

Table 1: Arguments passable to the gammamixEM function [14]

Argument	Required Y/N	Description	Default Value
x	Y	A vector containing the raw data used to fit the gamma distribution(s)	None- Requires user Input
lambda	N	Initial value of the mixing proportions	Selected from random uniform Dirichlet distribution
alpha	N	Initial vector of values of the gamma function shape parameter	Estimated by maximum likelihood method from lambda
beta	N	Initial vector of values of the gamma function scale parameter	Estimated by maximum likelihood method from lambda
k	Y	Number of gamma distributions to fit	None- Requires user input
epsilon	N	Convergence criterion	1×10^{-8}
maxit	N	Maximum number of before terminating due to non-convergence	1000
maxrestarts	N	Maximum number of times the function can restart the iteration counter and start with new initial values if the function determines that it is not converging quickly enough	20
verb	N	A setting to give verbal updates after each iteration of the function	Logical FALSE

The *lambda*, *alpha*, and *beta* parameters are derived by random number generation but are not generated independently. If no input for *lambda* is given, the function will partition the data into a number of regions equal to the given *k* value by drawing *k* numbers from a random uniform Dirichlet distribution. This means that the random uniform numbers drawn will sum to one. Then *alpha* and *beta* parameters will be estimated based on the *lambda* proportions by a method-of-moments estimate [14].

2.1.2 Global Optimization Issue with High Number of Distributions

As previously mentioned, if one does not approximate and specify all of the inputs into the gammamixEM function, the result will have increased variability. This will be especially true when the *lambda*, *alpha*, and *beta* parameters are not specified, as their starting point will be derived randomly under this scenario. Figure 5 – 10 below shows an example of different outputs using the same input data fitting gamma distributions using only the default parameter inputs. Figure 5 shows three outputs fitting a single distribution, Figure 6 shows three outputs fitting two distributions, Figure 7 shows three outputs fitting three distributions, Figure 8 shows three outputs fitting four distributions, Figure 9 shows three outputs fitting five distributions, and Figure 10 shows three outputs fitting six distributions. It can be seen from this example that providing no changes to the default parameter inputs causes variability of the output but as one fits a higher number of distributions the variability also increases.

Initially it was believed that this issue could be corrected by reducing the convergence criterion and increasing the number of iterations. By doing this the function would need to obtain a lower log-likelihood difference between iterations in order to converge on an optimum solution and the function would have more iterations in which to reach this convergence threshold. However, it was found that the function would typically further optimize a local optimum instead of moving toward a global optimum so this approach was abandoned.

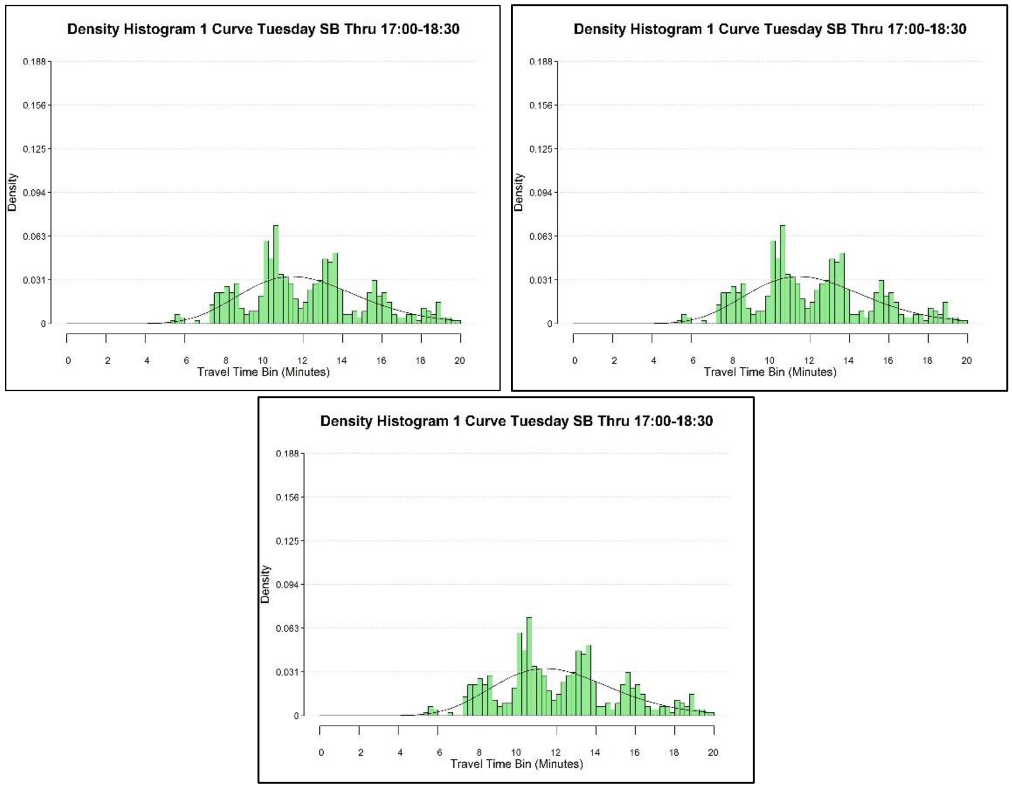


Figure 5: Density histograms of fitted single gamma distributions comparing variability among replicate fitting trials

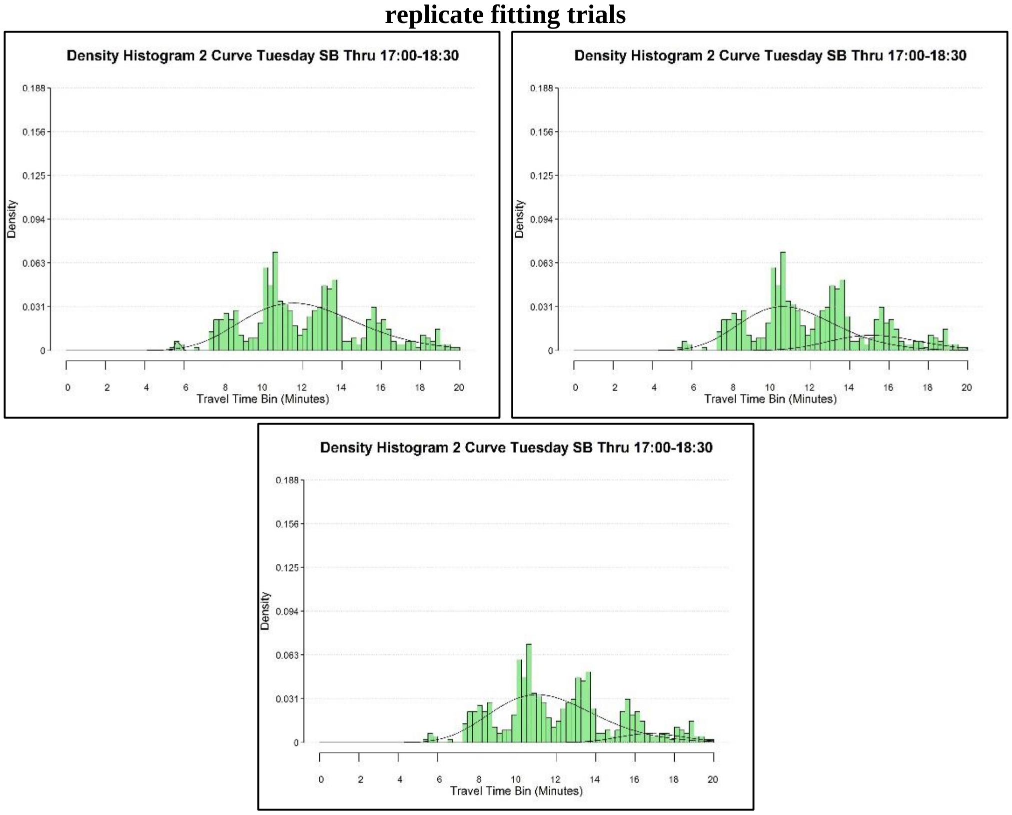


Figure 6: Density histograms of fitted 2 gamma distributions comparing variability among replicate fitting trials

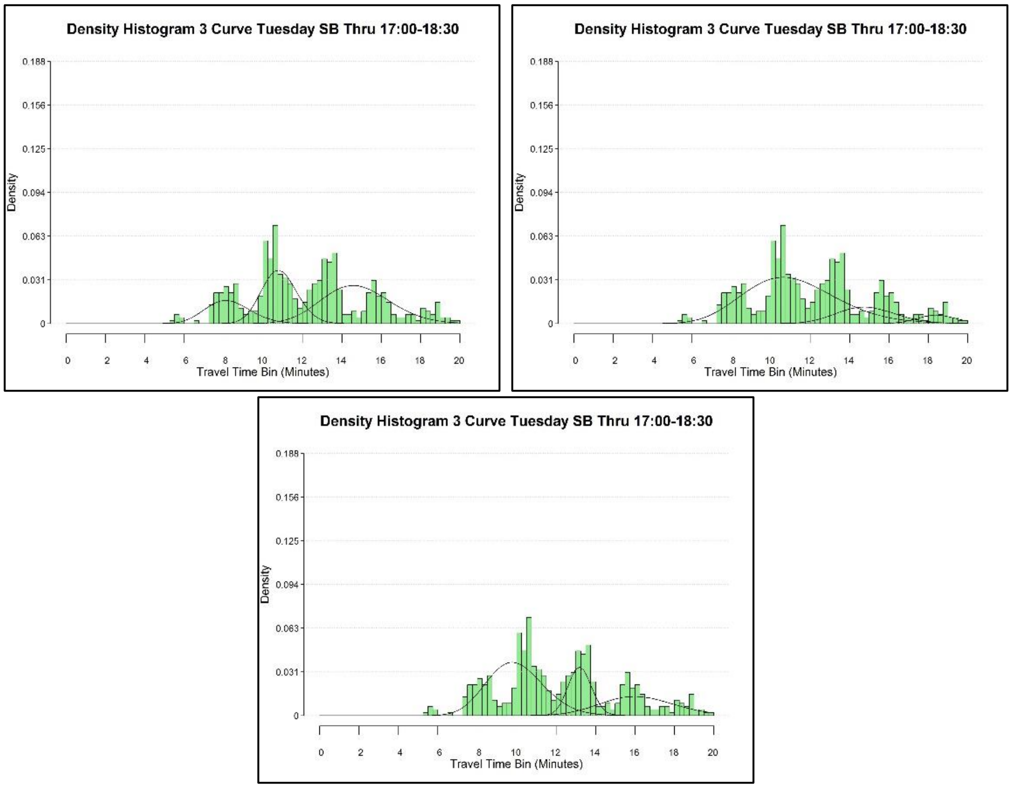


Figure 7: Density histograms of fitted 3 gamma distributions comparing variability among replicate fitting trials

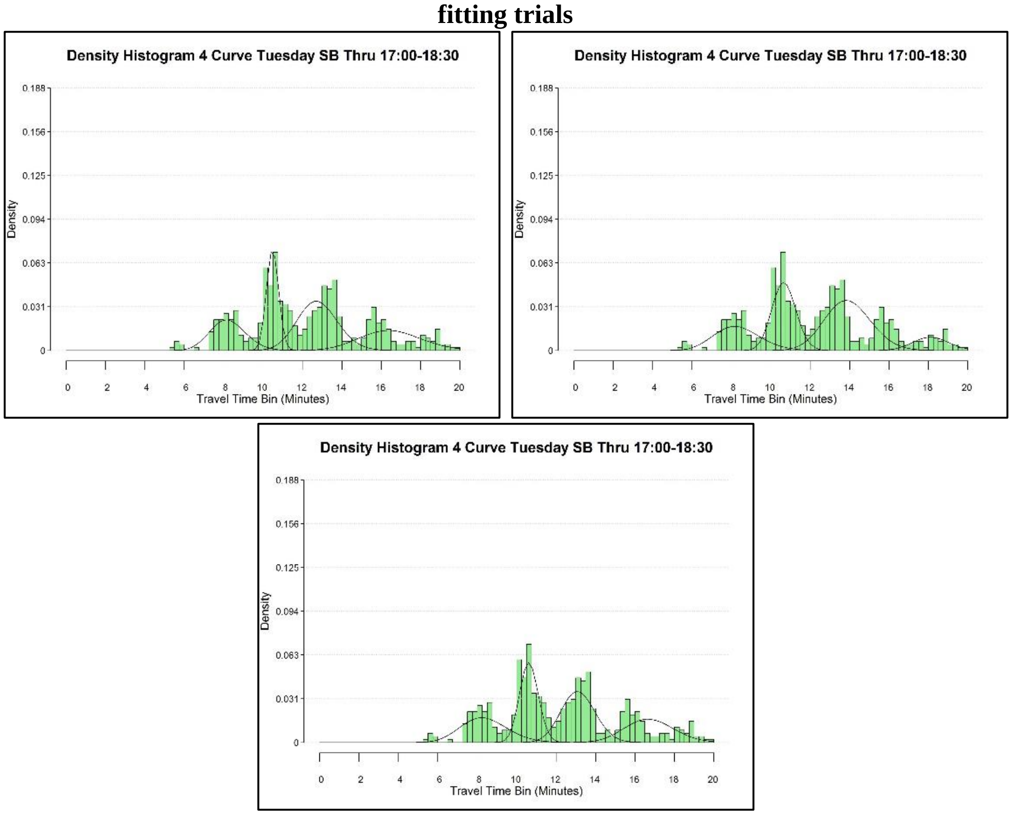


Figure 8: Density histograms of fitted 4 gamma distributions comparing variability among replicate fitting trials

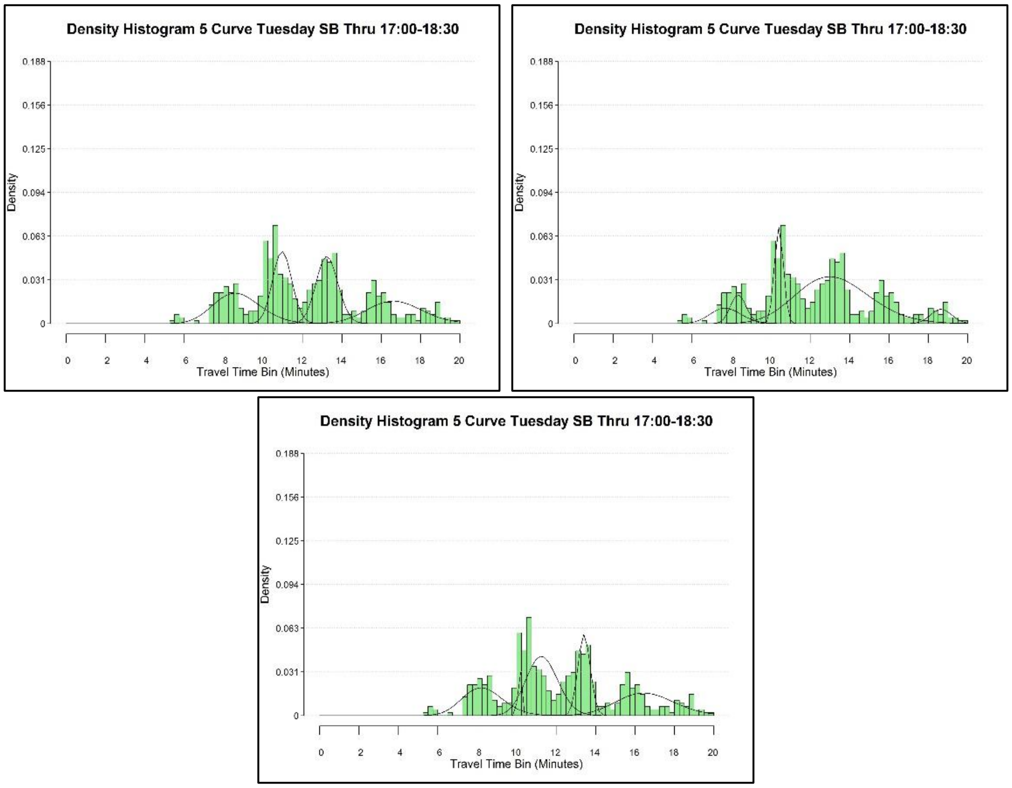


Figure 9: Density histograms of fitted 5 gamma distributions comparing variability among replicate fitting trials

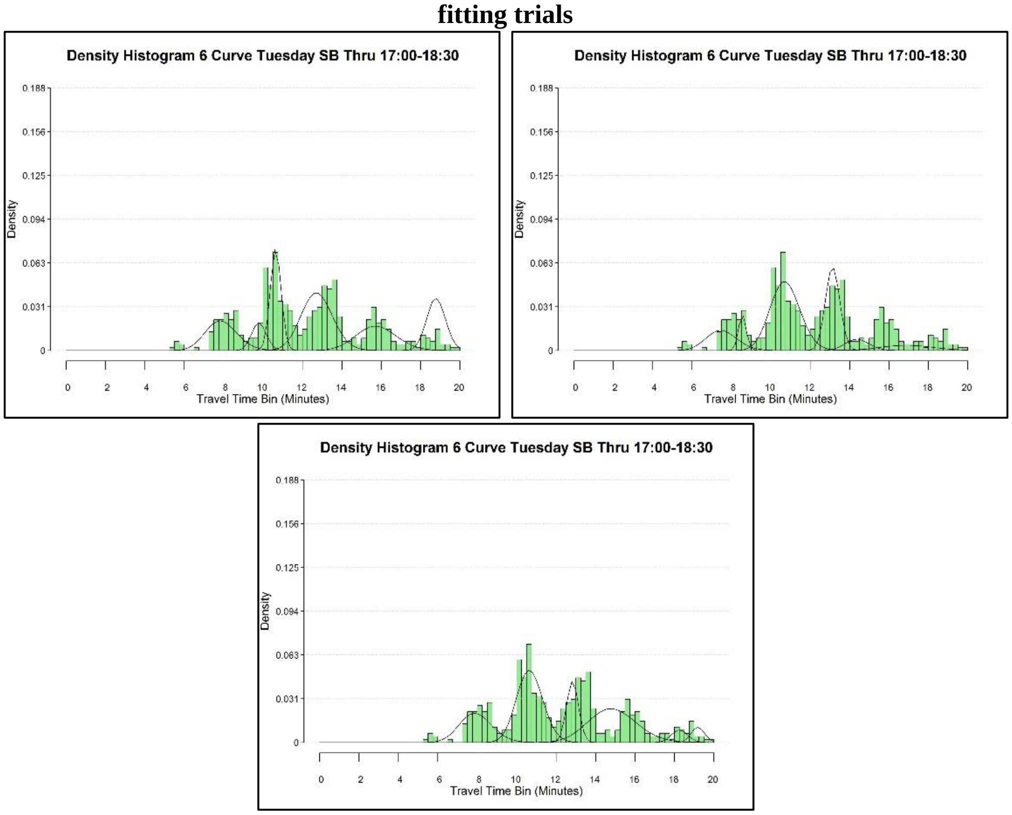


Figure 10: Density histograms of fitted 6 gamma distributions comparing variability among replicate fitting trials

Observing Figure 10 – 10, it can be seen that when fitting only a single gamma distribution the resulting curve appears to be the same for each of the three trials. This shows that the function is capable of optimizing to a single global optimum without needing to specify specific information for the non-required inputs. Next, comparing the two gamma distribution fits it can be seen that two of the fits appear similar while one shows a very different solution. This shows that at two distributions out of three runs the function converged on two different optimum, at this level it is difficult to determine which, if any, of the two optimum are a global optimum. Next, observing the three gamma distribution fits, the function converged on three separate optimums. Again, it is difficult to determine without further analysis which, if any, are a global optimum. The result test for four, five, and six distributions is similar to that of the three distribution, in that three separate solutions were found and further analysis is needed in order to determine which, if any, are a global optimum.

2.1.3 Global Optimization by Monte Carlo Trials

The Jimmy Carter Boulevard DDI before and after analysis has many sets of travel time data to which to fit gamma curves. There are ten routes, across seven time periods, over five days of data for a total of 350 data sets to which to fit gamma curves (10 routes * 7 time periods * 5 days = 350 data sets). Due to the large amount of data processing that would need to be done, it was desired to develop a method to find a global optimum without estimating the non-required inputs before fitting each set of curves. Furthermore, it was also desired to develop an automated method to find which number of distributions best describes the data.

First, the global optimization issue had to be solved in order to ensure that when determining the number of distributions that best describes the data comparisons are made between globally optimized gamma distributions. In order to accomplish this a Monte Carlo approach was taken. In this approach, each dataset was fit 100 times for each number of gamma distributions under consideration (1 to 6 gamma distributions) excluding a single gamma distribution, as at the single distribution level the function is very stable. This makes 507 runs for each dataset, for a total of 177,745 total fits (507 runs per dataset * 10 routes * 7 time periods per day * 5 days). Each single distribution fit was taken as a global optimum after a single run, however, for distribution fits one through six the best fit was used out of each of the 100 runs. The random variability is applied by allowing the function to draw the lambda parameter from a random uniform Dirichlet distribution on each run. After the completion of each run, an R-squared value was calculated. The first run is initially assumed to be the optimum and its output saved. Then after each successive run the R-squared value is compared to the one assumed to be the maximum, if the R-squared value on the current run is greater than the previous maximum then the new optimum function output is saved and compared to the next successive runs. If the current R-squared value is lower than the assumed max R-squared the R-squared value is saved, however, the function outputs are discarded. In this process, it is assumed that after 100 runs, the fit with the maximum R-squared will be a global optimum solution. Although it is recognized that this is not a guaranteed global optima and future effort is needed to develop a proven optima.

2.2 Identification of the Number of Distributions in a Dataset

After using a Monte Carlo method to globally optimize (assumed, as stated above) each set of gamma distribution fits, the optimized solutions for distribution fits from one to six were compared to each other to find which number of gamma distributions best describe the data. Previously, an R-squared value was used in order to determine the best fit among a group of fits using the same number of distributions. However, the R-squared value cannot be used to compare fits with a different number of distributions. This is due to the fact that as one adds more distributions to the mix, one is adding more variables with which to calibrate the fit, and thus reducing the degrees of freedom of the overall fit. This must be accounted for when comparing fits with different numbers of distributions, as reducing the degrees of freedom affects the information content of the fit by lowering the functions predictive capability. Therefore, finding the correct number of distributions is a balance between goodness of fit of each of the individual distributions and the degrees of freedom of the overall fit. In order to account for both of these considerations, the Akaike Information Criterion (AIC) was used to determine the best fitting number of distributions.

The AIC value is calculated by the equation in Figure 11. The gammamixEM function automatically computes a final log likelihood for the fit and this log likelihood was used in the AIC equation to compute the log(L) term. The k term in the equation in Figure 11 is the number of parameters used to fit the model. For a gamma distribution four parameters are used and thus k increases by four for each addition to the number of distributions in the model. Using this criterion, the more parameters that are added to the model the higher the AIC. Thus, when determining the number of distributions that best describes the data one would choose the model that has the lowest AIC [15].

$$AIC = -2 * \log(L) + 2 * k$$

Figure 11: Akaike's Information Criterion equation [15]

2.3 Separation of Data into Distributions

The goal of this entire process is to create a statistical method to separate the travel-time data into multiple distributions. Data separation actually takes place in this process before calculating the R-squared value for each of the 100 gamma distribution fits described in Section 2.1.2 above as the separated data are used to calculate R-square. However, the process was not described and will be provided here.

After fitting multiple distributions to each data set, the function provides a posterior probability for each travel time data point for each distribution. This means that if four distributions were fit to the dataset, then each data point will have four posterior probabilities assigned to it, the sum of which equals one. These posterior probabilities are the probabilities that each data point belongs to a particular distribution. Because these are only probabilities they are not used to directly assign each point to a distribution by choosing to assign each point to a distribution based on the highest probability. Instead, they are used in combination with a random uniform number between 0 and 1 to assign each data point to a distribution. The following methodology is repeated for each data point under consideration.

2.3.1 Data Separation Algorithm

1. A random uniform number between 0 and 1 is generated using a random uniform number generator in R.
2. The random uniform number generated in step 1 is iteratively compared to the data point under consideration's first posterior probability, then the sum of the first and second posterior probability, through to the sum of all the posterior probabilities checking at each iteration the criteria stated in step 2a and 2b
    a. If the random uniform number is less than the current sum of posterior probabilities then assign this data point to the current iteration's distribution and move to the next data point starting with step one (e.g. if one is checking against the sum of the first, second, and third posterior probabilities then assign the point to distribution three).
    b. If the random number is greater than the current sum of posterior probabilities then move to the next iteration of step 2.
3. Repeat steps 1 and 2 until all data points have been assigned to a distribution.

CHAPTER 3

RESULTS AND DISCUSSION

This section will describe some assumptions in the pre-filtering methods and results specifically for a single day of the week from each of the 10 routes. Since Tuesday is considered to be a fairly average travel day in terms of work trips, school trips, and other miscellaneous trips (e.g. shopping, dining, or other activities), the combined data over the 13 Tuesdays within this study will be shown with respective final multiple distributional fits. Generally, data were not combined across days (e.g. mixing data from Monday and Tuesday together) because it was believed that different days of the week typically see different levels of traffic congestion. This becomes apparent when you look at plots of combined data from different days. Typically, it appears that Thursday and Friday seem to experience higher levels of congestion than Monday, Tuesday, or Wednesday. Observing Figure 12, one can see 5 plots of travel time data for the route from Best Friend Road to Rockbridge Road (SB red line in Figure 2a). Each plot is of combined data for a different day of the week (e.g. Monday, Tuesday, Wednesday, etc.) during the same time period each day (3:00PM – 5:00PM). Looking through each plot one can see there are similar patterns early on in the time period, however, congestion sets in at different times each day, and on Thursday and Friday appears to rise to a higher level than other days of the week.

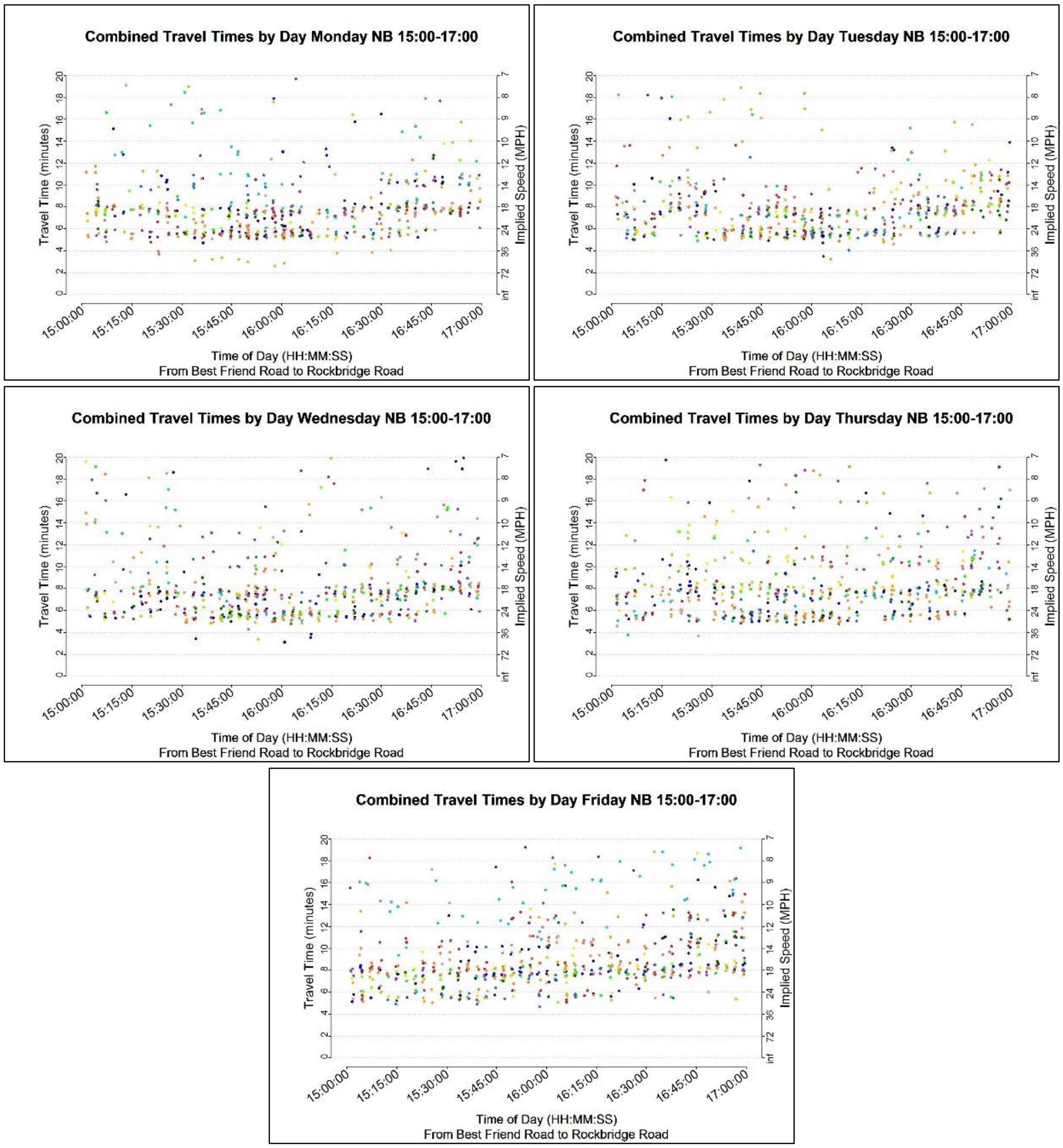


Figure 12: Travel time data plots from Best Friend Road to Rockbridge Road from 3:00PM to 5:00PM for each day of the week

3.1 Pre-Multiple Distribution Fit Filtering and Binning

Before fitting multiple distributions to the travel time data, a few filters were put in place to remove extreme data points that would clearly not belong to the distributions of vehicles that do not stop along the corridor. These filters are:

1. A 20-minute filter that removes and travel time data that is above 20 minutes, and 2. A 60 miles per hour (MPH) filter that removes travel times which have an implied speed above 60 MPH on the 45 mph speed limit roadway

Furthermore, the data were split into separate time periods by signal timing period in order to eliminate the possibility of multiple distributions occurring due to changes in signal timing.

3.1.1 20-Minute Filter

It is assumed that travel times above the 20 minute threshold fall into one of two cases. The first case is that the vehicles stopped along the corridor for a short period of time and then resumed travel along the corridor artificially inflating their travel time. The second case is one that can be encountered when using raw data matches. Here, the commercial Bluetooth® system used in this research finds MAC address matches between Bluetooth® detectors from a continuously updated list of MAC addresses from each detector. In order for this task to be manageable, the algorithm attempts to find matches between the two detectors' lists in approximately 1.5 hour windows (as inferred from the maximum observed travel time in the raw data files provided by the Bluetooth service provider). This means that if a vehicle carrying the same Bluetooth® device travels the corridor twice within a 1.5 hour window then it will be detected twice at each detector generating 4 travel times. Two of the travel times will be actual travel times and the other two travel times will be unwanted matches. One unwanted match will be a positive, and possibly believable, travel time and one will be a negative travel time. The match between the first detection at the first location and the second detection at the second location will be a positive, but virtual travel time. The match between the second detection at the first location and the first detection at the second location will generate a negative travel time, and is discarded by the commercial Bluetooth® system (as the data sets do not contain any negative travel times). This is illustrated in a generalized time vs. space sketch in Figure 13, where the green arrows represent the actual trips and red arrows represent unwanted matches.

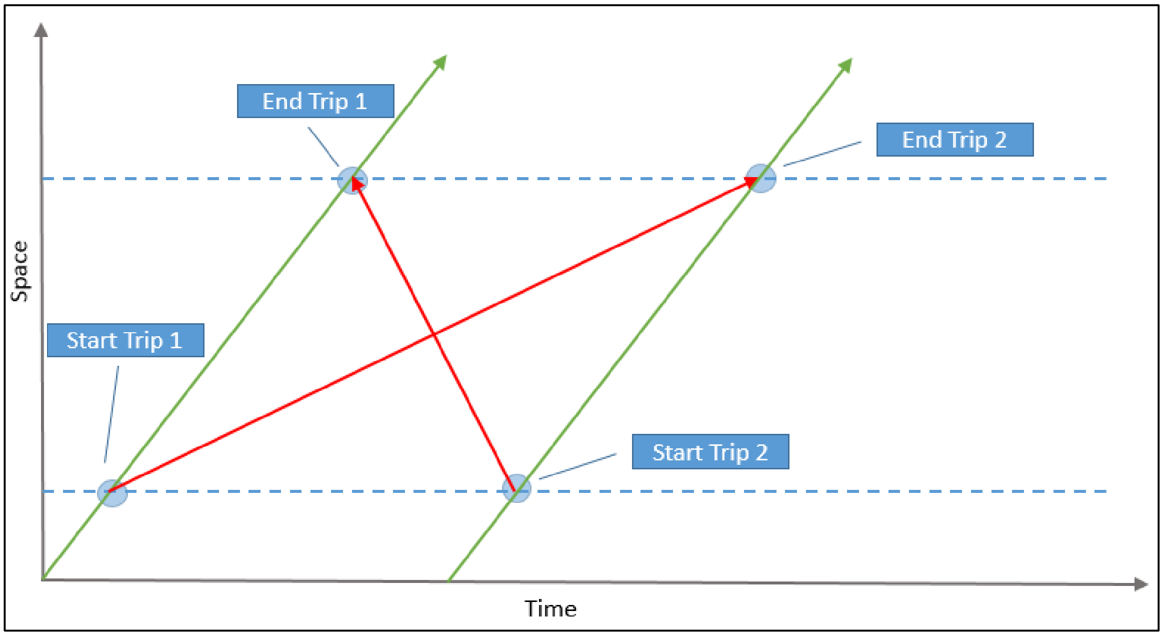


Figure 13: Time vs. space sketch of two trips generating 4 travel times

3.1.2 60 Miles Per Hour Filter

The 60 MPH filter is used to remove travel time data whose implied speed (calculated by dividing distance traveled by travel time) is above 60 MPH. Here it is believed that vehicles cannot easily travel over 60 MPH during normal daytime operations along the corridor and that any travel times showing this speed may be caused by the large detection range of the commercial Bluetooth® detectors. Without knowledge of the Bluetooth® vendor's proprietary data filtering algorithms it is not readily known what processes are in place to handle the detection location of individual MAC addresses. Therefore, it is difficult to describe how the large detection range may affect the reported speeds with the commercial detectors. Two methods are described next with their possible effects on reported travel times.

Two typical methods of handling the detection location of Bluetooth devices is to use a First-in/First-out (FIFO) or a Last-in/Last-out (LILO) scheme. In the FIFO scheme, the first detection occurrence of each Bluetooth® device at each location is used for matching between detectors. LILO is similar with FIFO except using the last detection occurrence at each location. Assuming a similar detection range of each detector the detections should occur at similar distances away from the detector. This also helps to reduce variability of the distance traveled between detectors leading to a safe assumption of the same distance travelled between detectors. Although, if one were using a FIFO scheme, and a Bluetooth® device is not detected for the first time until it is far downstream from the detector, and then the first detection at the second detector is far upstream from the detector the distance over which the vehicle traveled is much smaller than if the first and second detection occurred upstream of the detector. Here, one may calculate an implied speed that is much higher than the actual speed due to the reduced travel time reported and the assumption that all trips between the detectors are the same length. In short, the FIFO scheme has uncertainty in location due to the detection process and the signal strength while the LILO scheme only has uncertainty due to the signal strength. In order to reduce this possibility of error in travel time due to either a FIFO or LILO scheme, the 60 MPH speed filter was implemented.

3.1.3 Binning By Signal Timing Period

The RAW data from the Bluetooth® detection system were binned into seven different time periods for the analysis. These seven time periods were used because the signal timing plans on this corridor are changing throughout the day, primarily during the morning and afternoon peak periods, in order to accommodate typical traffic observed on the corridor during certain times of day. The signal timing plans are set to occur by time of day, and during our 13-hour analysis time period window of 6AM to 7PM, seven timing plans are used. In order to avoid the possibility of multiple distributions occurring due to changes in signal timing during a single analysis period, each of the timing plan periods were analyzed separately, creating seven time-period bins.

3.2 Tuesday Results

3.2.1 From Best Friend Road to Rockbridge Road

In Figure 14, one can see the density histograms with final fitted gamma curves for each of the seven time periods analyzed for the route from Best Friend Road to Rockbridge Road. This is the southbound route (red line in Figure 2a) from end to end of the corridor. This route is one of the most prominent displays of multiple distributions throughout the entire day. The lowest number of distributions to be determined for this route was four during the 7:45AM – 8:45AM period. The highest number of distributions determined for this route was six for both the 3:00PM – 5:00PM and the 5:00PM – 6:30PM time periods.

During the AM peak period from 6:00AM – 9:30AM multiple distributions are also observed along this route. In the first timing period from 6:00AM – 7:45AM congestion is relatively low and two distributions are very prominent while the last three carry less much fewer data points. However, during the 7:45AM – 8:45AM and during 8:45AM – 9:30AM periods congestion builds in this direction and four distributions become fairly prominent. However, the algorithm has grouped what appear to be two distributions together into one distribution in the 7:45AM – 8:45AM timing period. Upon reviewing the AIC values, it was found that four distributions had an AIC of 958 while the next closest at six distributions had and AIC of 974. It was also found that the R^2 output for four distributions was lower, at 0.856, than for six distributions, at 0.950. This suggests that according to the R^2 value six distributions is a better fit than four distributions, however, the extra parameters cause too much loss in the fits' predictive ability for six distributions to be a viable option.

During the midday period from 9:30AM – 3:00PM the corridor is relatively uncongested but still displays three prominent distributions with two distributions carrying a very low number of data points. This would suggest during this period the coordinated signal system may not be designed to fully optimize this direction of travel. This could be intentional, as fully optimizing this route may cause excessive delays on other routes. In any case, the gamma distribution fits appear to be very good here and in a before and after study the first two distributions would be the distributions of interest when observing any shifts in travel time.

Coincidentally, the PM peak period from 3:00PM – 7:00PM is when congestion in this direction is heaviest as drivers are travelling home from work or travelling for miscellaneous trips. Here the algorithm worked very well again matching each prominent distribution closely and providing a sound method of separating the distributions.

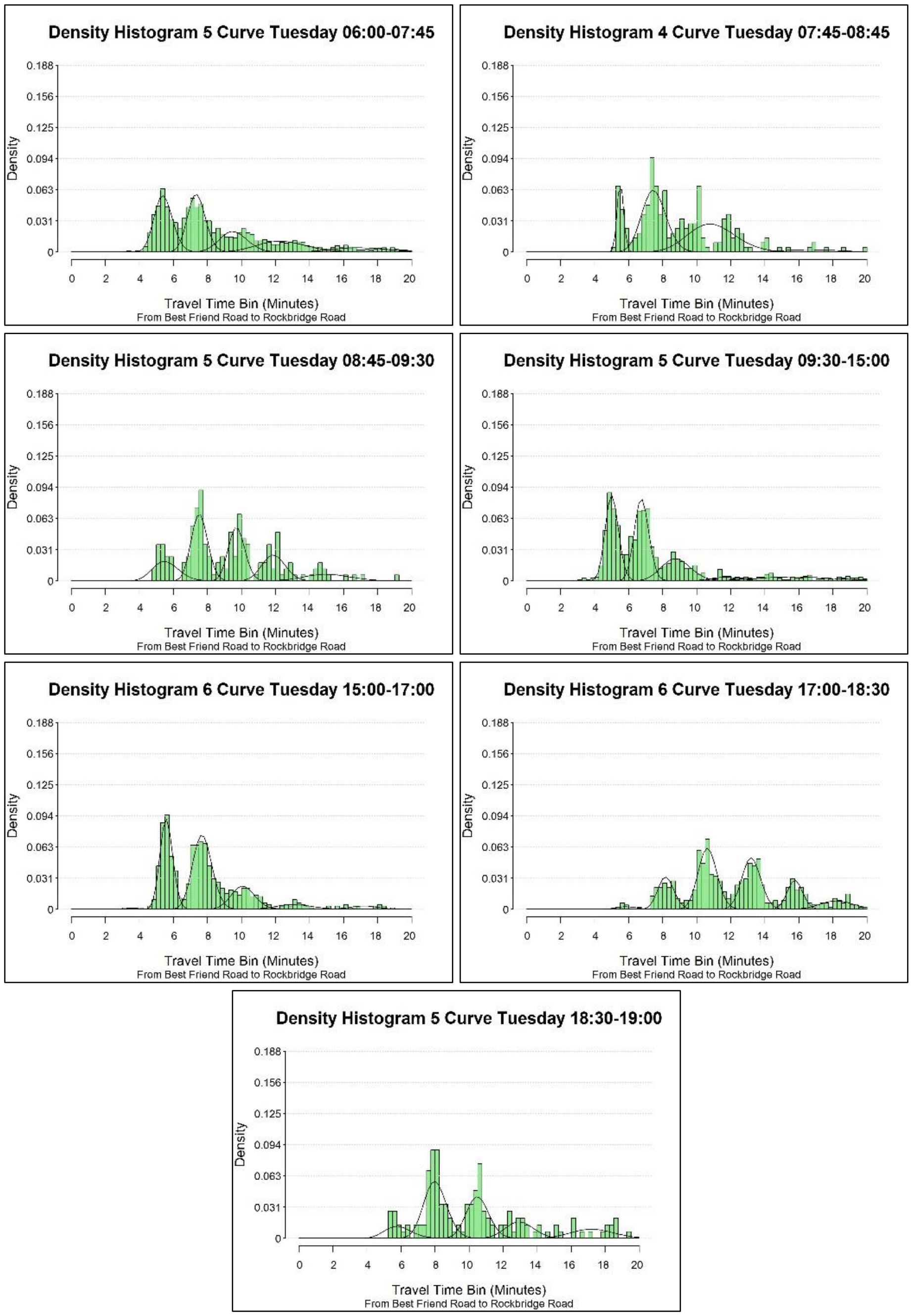


Figure 14: Multi-gamma distribution fit histograms for Best Friend Road to Rockbridge Road Route

3.2.2 From Rockbridge Road to Best Friend Road

Below in Figure 15 one can see the density histograms with final fitted gamma curves for each of the seven time periods analyzed for the route from Rockbridge Road to Best Friend Road. This is the northbound route (green line in Figure 2a) from end to end of the corridor. The lowest number of distributions to be determined for this route was a single during the 7:45AM – 8:45AM period. The highest number of distributions determined for this route was three for the 9:30AM – 3:00PM time period.

During the AM peak period from 6:00AM to 9:30AM there are two signal timing periods that were fit with two distributions (6:00AM – 7:45AM and 8:45AM – 9:30AM) and one was fit with a single distribution (7:45AM – 8:45AM). The two distribution fits are interesting because the bulk of the data points were fit into the first distribution while very few were fit into the second distribution. It is possible that the data points contained in the second distribution could be considered outliers that were either vehicles that stopped briefly along the corridor or traveled the corridor twice. During the 7:45AM – 8:45AM time period congestion is more intense than the other two time periods which reduces the likelihood that data points in the extreme near 20 minutes are in fact outliers. Therefore, during this time a single distribution better describes the data.

Surprisingly, during the midday period from 9:30AM to 3:00PM, three distributions were determined to be the best fit. Observing the data itself, it appears that two distributions or a single distribution is a more likely best fit. Observing the AIC values, the three distribution fit had a value of 4181.3 while a two distribution fit had a value of 4201.21. This suggests statistically it makes more sense to fit three distributions, however, logically it may make more sense to choose two distributions since the goal of the analysis is not to break up obvious distributions from over-fitting, but rather to isolate the distributions for further analysis in a before and after study.

During the PM peak period from 3:00PM – 7:00PM one and two distributions were determined to be the best fit. Similar to the AM peak period, the data in the first distribution is of utmost interest and data contained in the second distribution (where a second distribution exists) may be considered to be outliers.

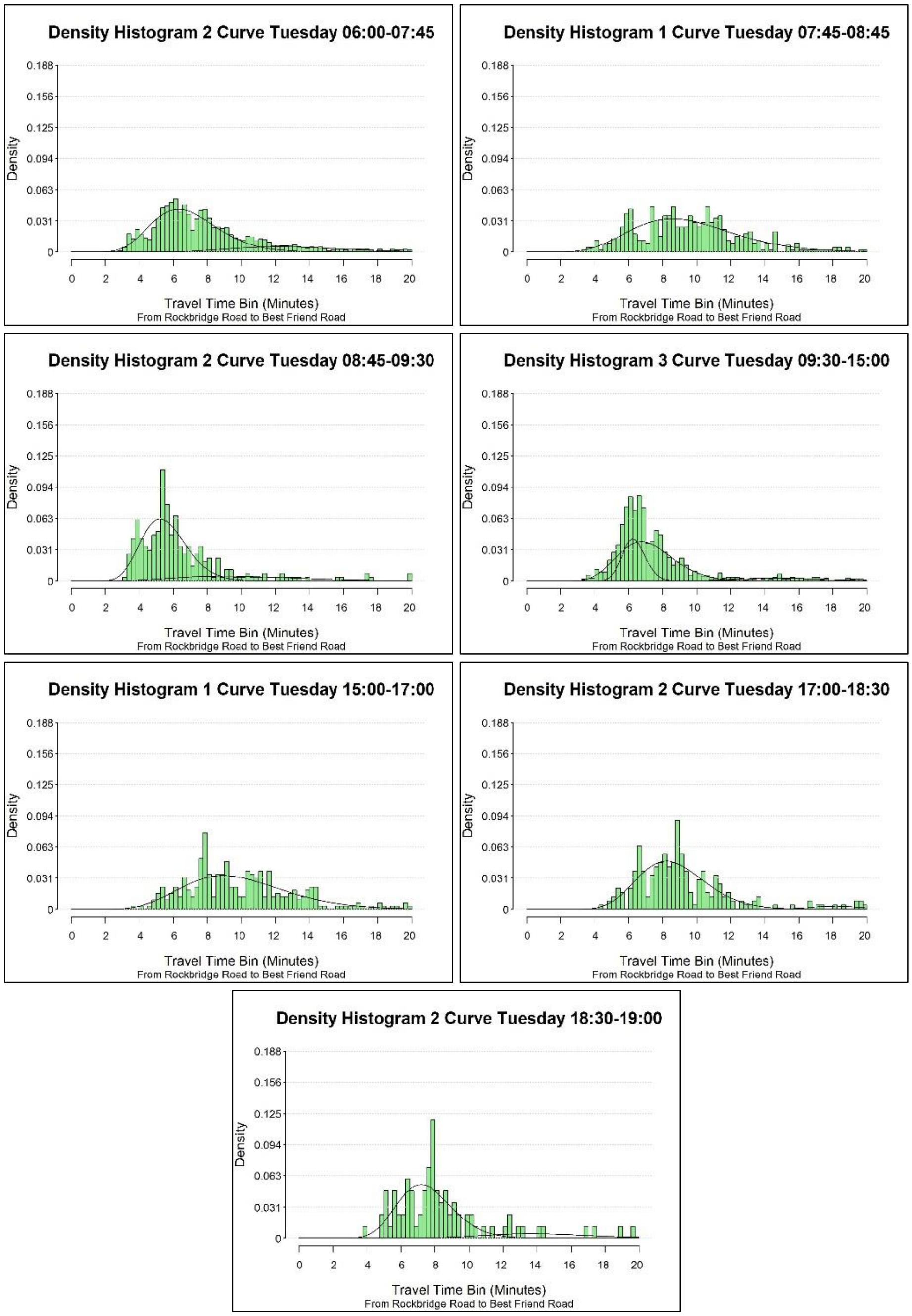


Figure 15: Multi-gamma distribution fit histograms for Rockbridge Road to Best Friend Road Route

<u>3.2.3 From Best Friend Road to I-85SB on ramp</u>

Below in Figure 16 one can see the density histograms with final fitted gamma curves for each of the seven time periods analyzed for the route from Best Friend Road to the I-85SB on ramp. This is the southbound route (red line in Figure 2b) starting at Best Friend Road and making a right turn onto I-85SB. The lowest number of distributions to be determined for this route was a single distribution during the 5:00PM – 6:30PM period. The highest number of distributions determined for this route was six for both the 6:00AM – 7:45AM time period and the 9:30AM – 3:00PM time period.

During the AM peak period from 6:00AM – 9:30AM the two solutions during the 7:45AM – 8:45AM and 8:45AM – 9:30AM time periods appear to be good fits to the data. In these data sets, the first distribution is of most interest as the other two may be considered to contain outlier data. However, in the 6:00AM – 7:45AM time period six distributions were fit to the data which clearly may contain fewer distributions. In this instance, the second best fit according to the AIC values was five distributions at 15413 while six distributions had an AIC value of 15372. This is possibly another case where statistically it makes more sense to fit six distributions but logically one would desire to use the five distribution fit in order to keep from breaking up a single distribution for analysis purposes.

During the midday peak period from 9:30AM – 3:00PM six distributions were also determined to be the best fit. However, this may also be a case where one would desire to use fewer distributions in order to keep from breaking up obvious individual distributions. In this case the five distribution fit was also the second lowest AIC value at 6262, while the six distribution fit was 6235.

During the PM peak period from 3:00PM – 7:00PM a variation of four distributions (3:00PM – 5:00PM), a single distribution (5:00PM – 6:30PM), and three distributions (6:30PM – 7:00PM) were found as optimal solutions. These distributional fits appear to do a good job describing the data. During the 3:00PM – 5:00PM time period the first two distributions would be of utmost interest as they carry most of the data, while the second two distributions likely carry a high number of outliers. It is interesting to observe that as congestion worsens around 5:00PM, and travel times begin to increase, a single distribution becomes the best fit capturing the increase in travel times through this time period. From 6:30PM – 7:00PM three distributions appear to capture the data very well and the third distribution appears to contain outlying.

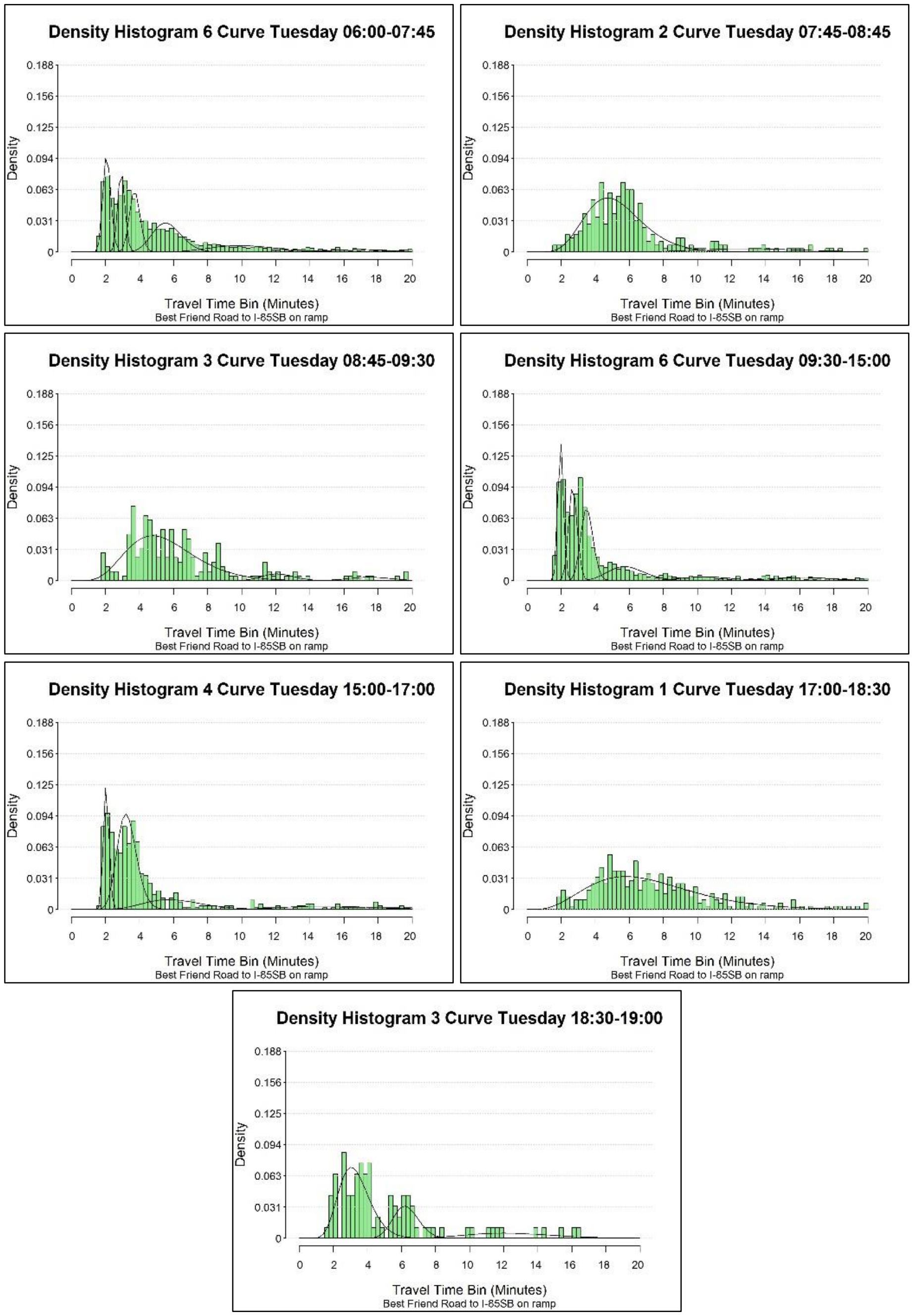


Figure 16: Multi-gamma distribution fit histograms for Best Friend Road to I-85SB on ramp

3.2.4 From Best Friend Road to I-85NB on ramp

Below in Figure 17 one can see the density histograms with final fitted gamma curves for each of the seven time periods analyzed for the route from Best Friend Road to the I-85NB on ramp. This is the southbound route (red line in Figure 2c) starting at Best Friend Road and making a left turn onto I-85NB. The lowest number of distributions to be determined for this route was a single distribution during the 5:00PM – 6:30PM period. The highest number of distributions determined for this route was five for the 6:00AM – 7:45AM, 7:45AM – 8:45AM, 9:30 – 3:00PM, and the 6:30PM – 7:00PM time periods.

During the AM peak period from 6:00AM – 9:30AM each of the three multiple distribution fits appear to be good fits. Again during this time period the last distribution in each of the fits most likely contains outlier data.

During the midday period from 9:30AM – 3:00PM five distributions were fit to the data set, where the last two data sets carry mostly data from the tail of the distribution and is most likely outlier data. Here, the algorithm found a very optimum solution at five distributions.

During the PM peak period from 3:00PM – 7:00PM each distributional fit appears to do a good job describing the data. As congestion builds during the 3:00PM – 5:00PM period and distributions become well mixed from the midday period the algorithm moved to using only 2 curves. Furthermore, when congestion worsens during the 5:00PM – 6:30PM period a single distribution is fit. Finally, as traffic is dissipating during the 6:30PM – 7:00PM the distributions again become a bit separated and five distributions are used.

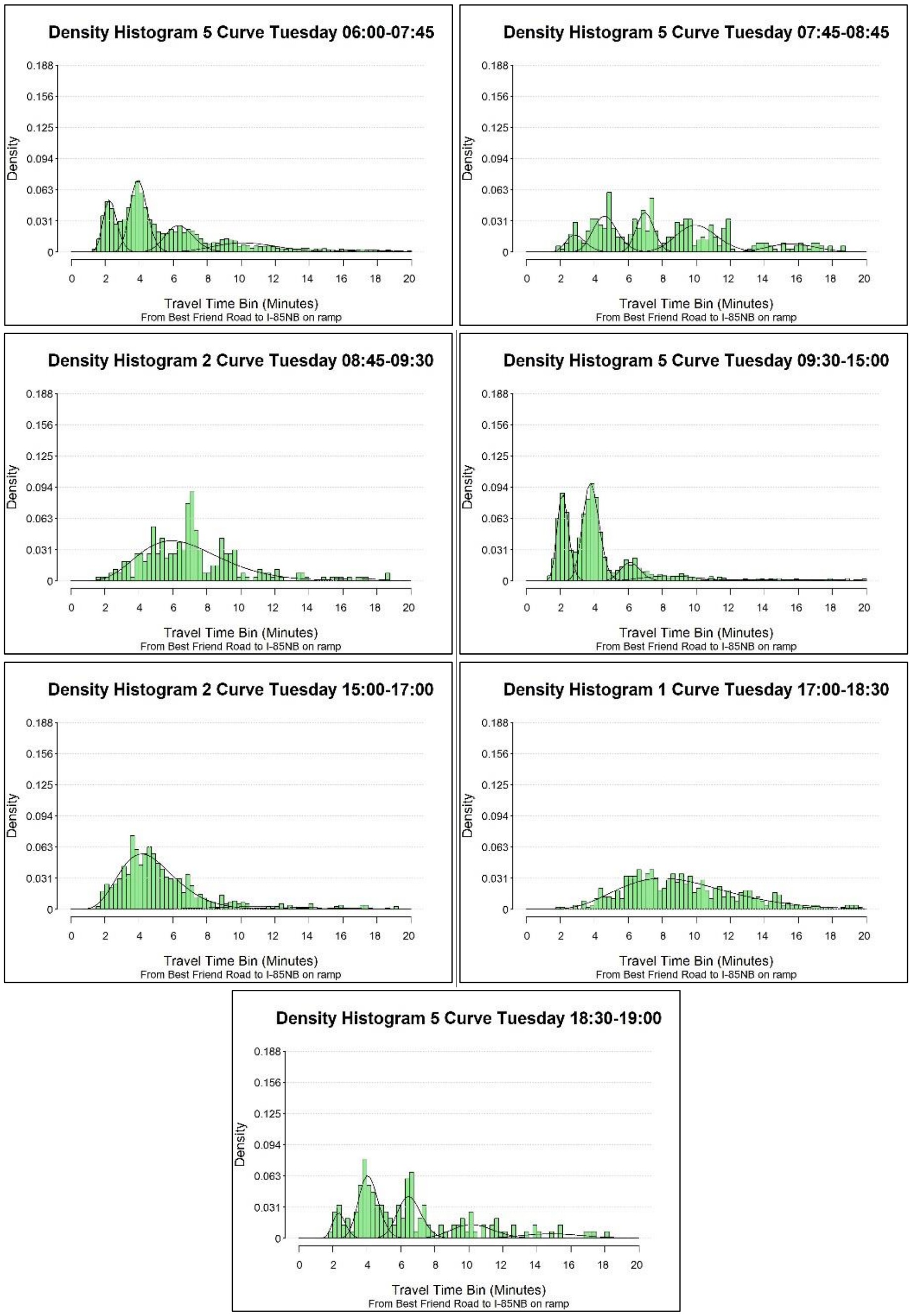


Figure 17: Multi-gamma distribution fit histograms for Best Friend Road to I-85NB on ramp

3.2.5 From Rockbridge Road to I-85SB on ramp

Below in Figure 18 one can see the density histograms with final fitted gamma curves for each of the seven time periods analyzed for the route from Rockbridge Road to the I-85SB on ramp. This is the northbound route (red line in Figure 2c) starting at Rockbridge Road and making a left turn onto I-85SB. The lowest number of distributions to be determined for this route was two during the 7:45AM – 8:45AM, 8:45AM – 9:30AM, 5:00PM – 6:30PM, and 6:30PM – 7:00PM time periods. The highest number of distributions determined for this route was three for the 6:00AM – 7:45AM, 9:30AM – 3:00AM, and 3:00PM – 5:00PM time periods.

During the AM peak period, it appears that travel times remain relatively uniformly spread with no or little break points between any possibly underlying distributions. Therefore, during this time two and three distributions were fit, with only the first distribution carrying what may be considered non-outlying data. It is interesting to note, that during the low congestion levels of the 6:00AM – 7:45AM time period three distributions were fit while the other AM time periods were fit with two distributions. It appears that when a very long tail is present, statistically, the algorithm finds a better solution by breaking the data in the tail into its own single or double distribution in order to produce a better fit for the larger portion of the data.

The midday period from 9:30AM – 3:00PM also found a similar three distribution solution as the early morning 6:00AM – 7:45AM solution. This solution is a very good fit to the data.

During the PM peak period from 3:00PM – 7:00PM two and three distribution solutions were found. The 3:00PM – 5:00PM time period has an interesting display of two overlapping distributions across the large peak in the data. This case may be breaking up the large peak that may be modeled better as a single distribution which is not ideal. Further inspection of the separated data will be needed in order to determine what level of impact this has on breaking up the main distribution. During the 5:00PM – 6:30PM and 6:30PM – 7:00PM time periods two distributions were fit and this appears to do a good job describing the data.

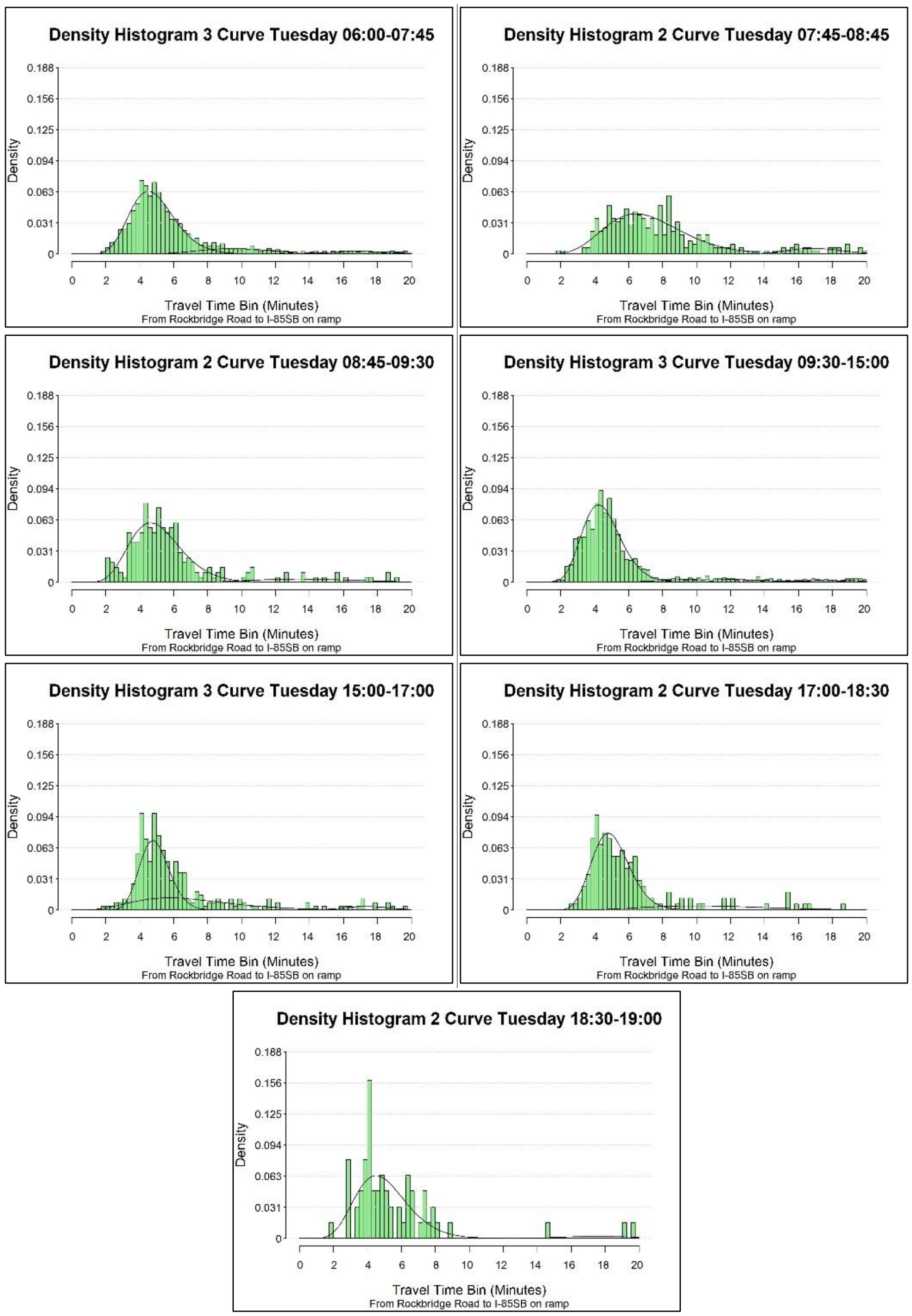


Figure 18: Multi-gamma distribution fit histograms for Rockbridge Road to I-85SB on ramp

3.2.6 From Rockbridge Road to I-85NB on ramp

Below in Figure 19 one can see the density histograms with final fitted gamma curves for each of the seven time periods analyzed for the route from Rockbridge Road to the I-85NB on ramp. This is the northbound route (red line in Figure 2b) starting at Rockbridge Road and making a right turn onto I-85NB. The lowest number of distributions to be determined for this route was two during the 8:45AM – 9:30AM time period. The highest number of distributions determined for this route was six for the 3:00PM – 5:00PM time period.

During the AM peak period from 6:00AM – 9:30AM four (6:00AM – 7:45AM), five (7:45AM – 8:45AM), and two (8:45AM – 9:30AM) distributions were fit. It can be seen during the earlier, less congested, time period from 6:00AM – 7:45AM four distributions were fit, which does a good job describing the data. Next, during the more congested 7:45AM – 8:45AM time period, five distributions were fit to the data, which also does a good job describing the data. However, during the 8:45AM – 9:30AM time period only two distributions were fit to the data. Here, again it appears that there are two large peaks that have been grouped into a single distribution. One may expect here that a similar solution would have been found to the 6:00AM – 7:45AM time period. Observing the AIC values for 8:45AM – 9:30AM one finds that for two distributions that AIC value was 1081 while the four distribution solution was 1089. As these are very close, one may choose to use four distributions instead of only two in order to analyze the two large peaks separately.

During the midday time period from 9:30AM – 3:00PM a five distribution solution was found. This solution appears to do a good job describing the data and allows for separation of outliers in the upper two distributions while the lower three distribution can be used for before and after analysis.

During the PM peak period from 3:00PM – 7:00PM the distributions fit appear to do a good job describing the data. This route, during this time period suffers from some additional random variability due to heavy congestion along I-85NB. This freeway congestion often backs up into the interchange causing excess delay not caused by the operation of the actual corridor. Therefore, distributions here appear to be a bit more mixed than other routes, however, it still appears the algorithm was able to capture the underlying distributions.

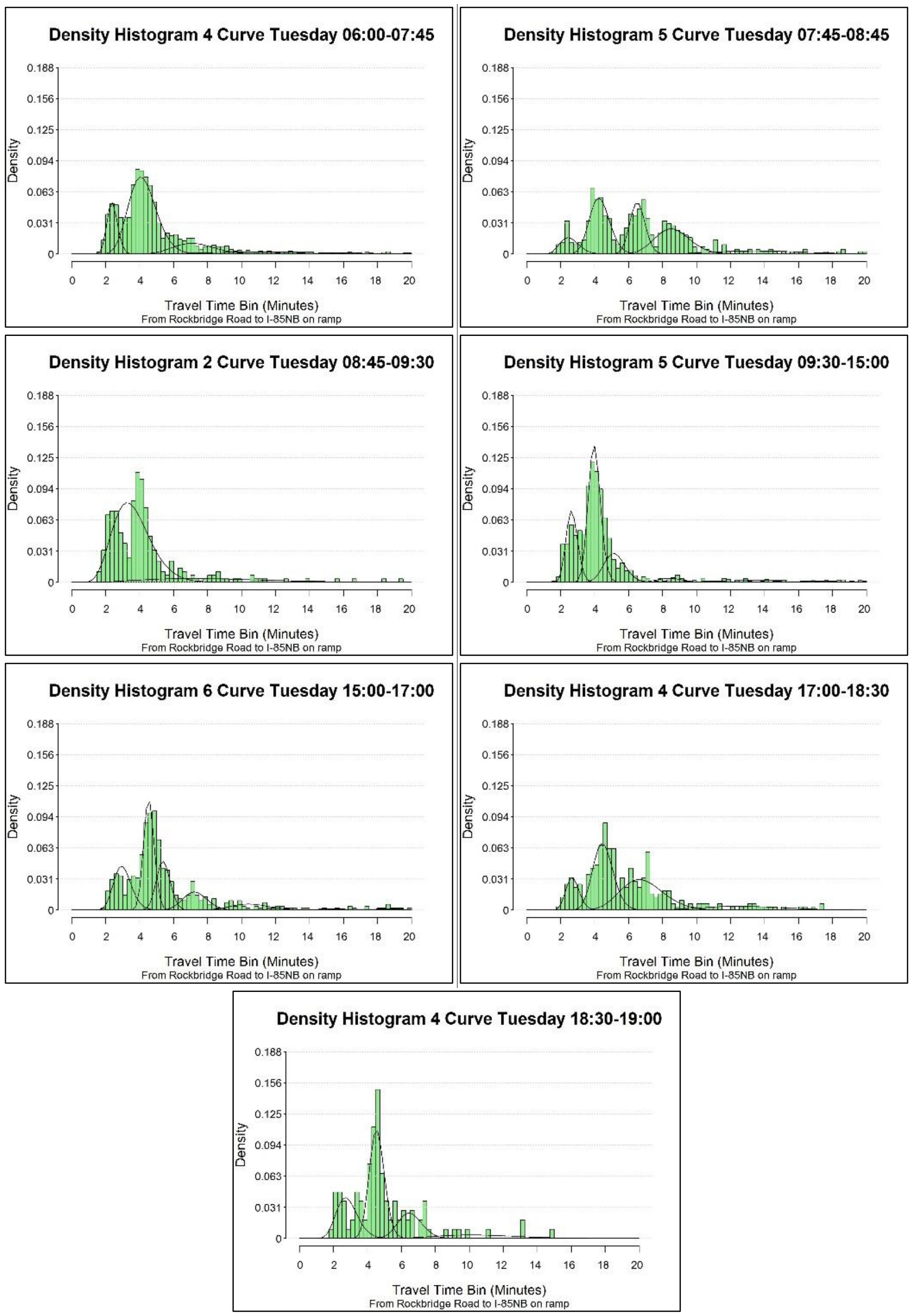


Figure 19: Multi-gamma distribution fit histograms for Rockbridge Road to I-85NB on ramp

3.2.7 From I-85SB off ramp to Best Friend Road

Below in Figure 20 one can see the density histograms with final fitted gamma curves for each of the seven time periods analyzed for the route from the I-85SB off ramp to Best Friend Road. This is the northbound route (green line in Figure 2d) starting at the I-85SB off ramp making a right turn onto Jimmy Carter Boulevard and ending at Best

Friend Road. The lowest number of distributions to be determined for this route was two during the 3:00PM – 5:00PM time period. The highest number of distributions determined for this route was four for the 8:45AM – 9:30AM and 9:30AM – 3:00PM time periods.

During the AM peak period from 6:00AM – 9:30AM three distributions were fit for both the 6:00AM – 7:45AM and 7:45AM – 8:45AM time periods while four distributions were fit during the 8:45AM – 9:30AM time period. All distributions fit to the data during this time period do a good job describing the data. During the first two time periods it appears that there is a single peak, fit to its own distributions, while the rest of the data were captured in the second two distributions and could be removed as outliers. From 8:45AM – 9:30AM it appears that the single large peak has broken up into two distributions, which was captured by the algorithm.

During the midday peak from 9:30AM – 3:00PM four distributions were fit to the data. Observing the histogram of the data, it appears that there is only one large peak distribution, however, two distributions were fit to this area while the other two were fit to outlying points. This case may have broken up the single peak into two distributions which is not ideal. During this time period it appears that a three distribution fit, similar to the 6:00AM – 7:45AM time period may logically be a better choice, in order to keep from breaking up single distributions. Here, the four distribution fit AIC value was 8419, while the three distribution fit was 8433.

During the PM peak from 3:00PM – 7:00PM, two distributions were fit to the data from 3:00PM – 5:00PM, and three distributions were fit to the data during the 5:00PM – 6:30PM and 6:30PM – 7:00PM time periods. The two distribution fit from the 3:00PM – 5:00PM time period describes the data well. However, during the 5:00PM – 6:30PM period three distributions were fit to the data, when logically it may be a better solution to use only two distributions. In fact, the AIC values between two and three distributions are very close at 1482 and 1482 respectfully. Furthermore, during the

5:30PM to 6:00PM time period, it may also make more logical sense to use only two distributions instead of three I order to keep from breaking up an obvious single distribution. Here, the AIC values for two and three distributions were 520 and 516 respectfully.

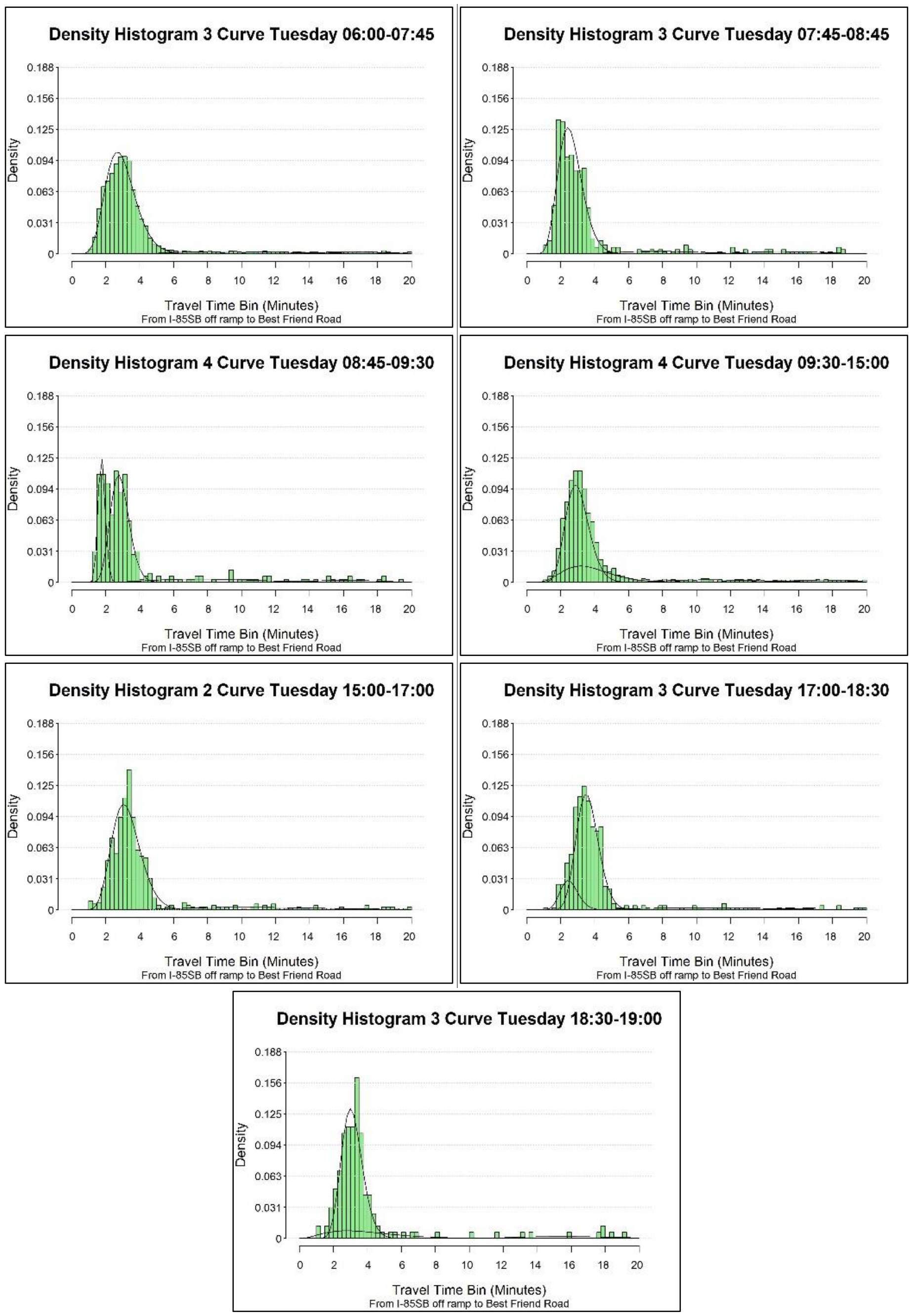


Figure 20: Multi-gamma distribution fit histograms for I-85SB off ramp to Best Friend Road

3.2.8 From I-85SB off ramp to Rockbridge Road

Below in Figure 21 one can see the density histograms with final fitted gamma curves for each of the seven time periods analyzed for the route from the I-85SB off ramp to Rockbridge Road. This is the southbound route (green line in Figure 2e) starting at the I-85SB off ramp making a left turn onto Jimmy Carter Boulevard and ending at Rockbridge Road. The lowest number of distributions to be determined for this route was three during the 6:00AM – 7:45AM, 7:45AM – 8:45AM, 8:45AM – 9:30AM, 9:30AM – 3:00PM, and 3:00PM – 5:00PM time periods. The highest number of distributions determined for this route was five for the 5:00PM – 6:30PM time period.

During the AM peak period, all three time periods were fit to three distributions. Generally, three distributions do a good job describing each of the data sets. During each of the time periods, there are a large number of travel times occurring at approximately 3.5 minutes with a much smaller collection of travel times occurring at approximately 5.5 minutes. Because of the nature of the gamma curves, and the small separation among these data it is difficult to fit two distributions to this data without splitting some of the data in the first distribution into the second distribution.

During the midday period from 9:30AM – 3:00PM, similar behavior is noticed in the travel time data as for the AM peak period. However, during this time three distributions were initially fit, but the two peaks (at 3.5 and 5.5 minutes) were later grouped into a single distribution. These data appear to be a bit more mixed than the AM peak data and the gammamixEM function may have difficulty discerning the two modes as separate distributions. An increased number of maximum iterations within the function, a stricter convergence criteria threshold, or both may help the function discern the two distributions.

During the PM peak period from 3:00PM – 7:00PM three (3:00PM – 5:00PM), five (5:00PM – 6:30PM), and four (6:30PM – 7:00PM) distributions were fit to the data.

Congestion on this route is increased during this time period and an additional third peak can be seen at about 8 minutes. The fits here appear to describe the data well.

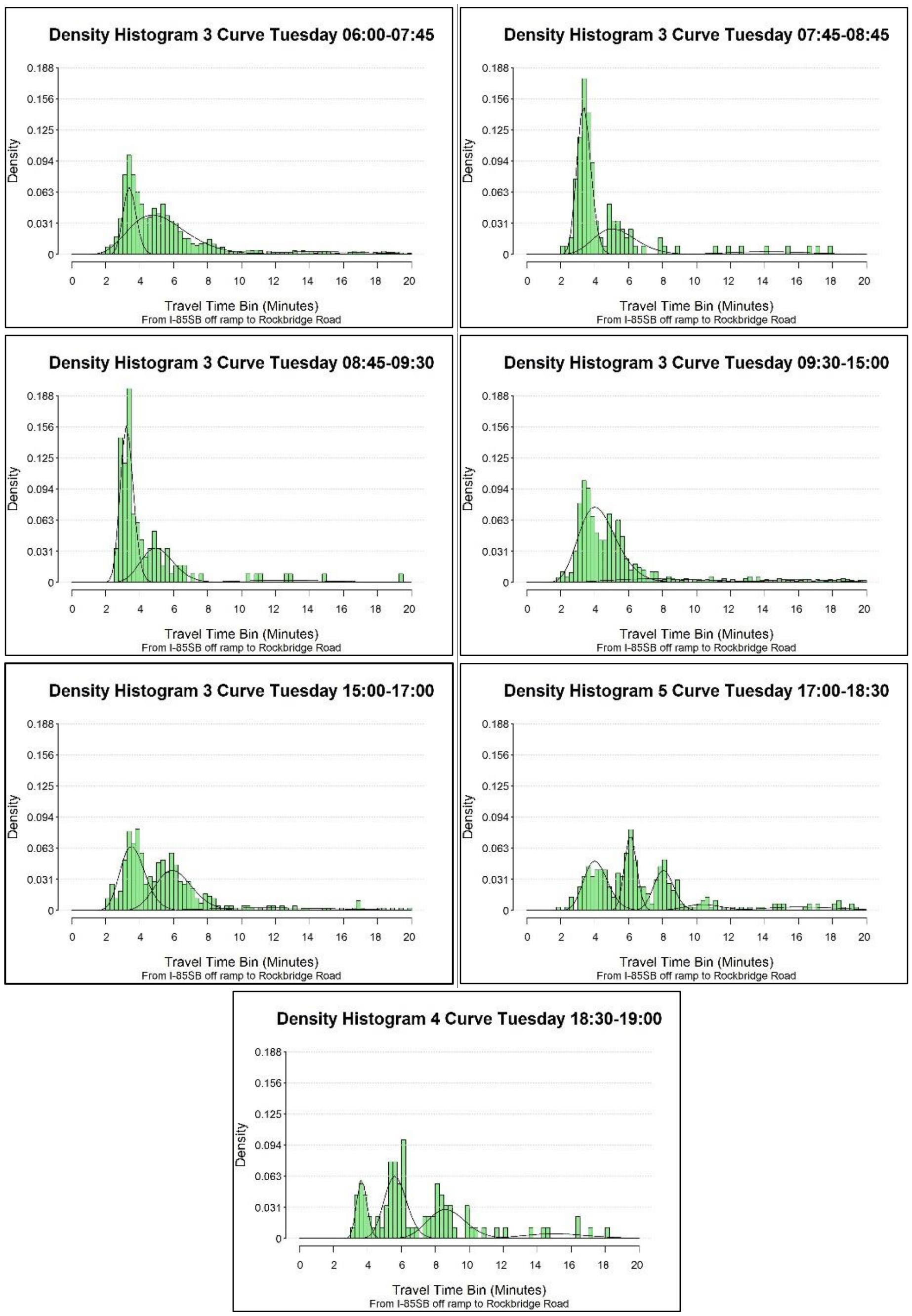


Figure 21: Multi-gamma distribution fit histograms for I-85SB off ramp to Rockbridge Road

3.2.9 From I-85NB off ramp to Best Friend Road

Figure 22 below presents the density histograms along with final fitted gamma curves for each of the seven time periods analyzed for the route from the I-85NB off ramp to Best Friend Road. This is the northbound route (red line in Figure 2e) starting at the I-85NB off ramp making a left turn onto Jimmy Carter Boulevard and ending at Best Friend Road. The lowest number of distributions to be determined for this route was two during the 3:00PM – 5:00PM, 5:00PM – 6:30PM, and 6:30PM – 7:00PM time periods. The highest number of distributions determined for this route was three for the 6:00AM – 7:45AM, 7:45AM – 8:45AM, and 8:45AM – 9:30AM time periods.

During the AM peak period from 6:00AM – 9:30AM three distributions were fit to the data during each of the three time periods. Three distributions appear to do a good job describing the data. In each of these the first distribution contains most of the data with the second two distribution capturing outlying data.

During the midday period from 9:30AM – 3:00PM three distributions were also fit to the data and describe the data well. Similarly to the AM peak period, the first distribution captures most of the data, while the second two distributions capture the remainder.

Interestingly, during the PM peak period each of the time periods analyzed were fit with only two distributions. This still describes the data very well, and similarly to the midday and AM peak periods the first distribution contains most of the data while the second distribution contains the outlying data.

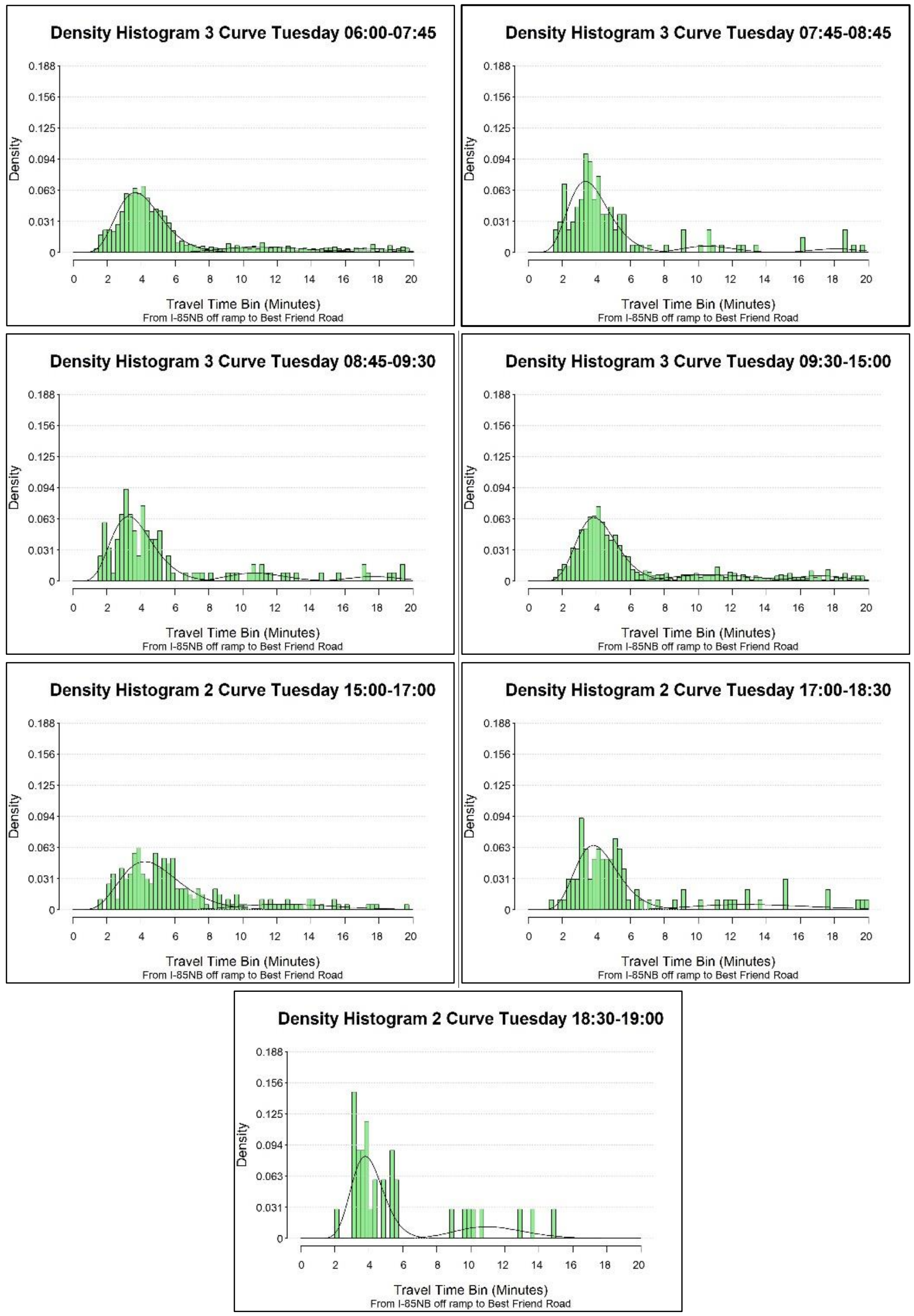


Figure 22: Multi-gamma distribution fit histograms for I-85NB off ramp to Best Friend Road

3.2.10 From I-85NB off ramp to Rockbridge Road

Figure 23 shows the density histograms with final fitted gamma curves for each of the seven time periods analyzed for the route from the I-85NB off ramp to Rockbridge Road. This is the southbound route (red line in Figure 2d) starting at the I-85NB off ramp making a right turn onto Jimmy Carter Boulevard and ending at Rockbridge Road. The lowest number of distributions to be determined for this route was two during the 6:30PM – 7:00PM time period. The highest number of distributions determined for this route was four for the 7:45AM – 8:45AM, and 8:45AM – 9:30AM time periods.

During the AM peak period from 6:00AM – 9:30AM three (6:00AM – 7:45AM) and four (7:45AM – 8:45AM and 8:45AM – 9:30AM) distributions were fit to the data. The data set from 6:00AM – 7:45AM appears to contain only a single prominent distribution, however, the algorithm appears to have broken it into two data sets. Further analysis will be needed to check if this is occurring. From the 7:45AM – 8:45AM time period, four distributions were chosen, which appears to do a good job describing the data, however, the first and second distribution do appear to have a strong overlap and will need to be investigated further for any breaking up of the first distribution. During the 8:45AM – 9:30AM time period, four distributions was chosen as the optimum, which describes the data well.

During the midday period from 9:30AM to 3:00PM, three distributions were chosen as the optimum. However, the first two distributions are overlapping. There, may be two underlying distributions here as can be seen in the slight separation at approximately 4.25 minutes. Therefore, this may be a good fit, however, it may be breaking up the first distribution which is not desirable. It was found that the AIC value for three distributions here was 3250 while the next best AIC value was 3269 for four distributions. In this case it may make more logical sense to go with four distributions in order to reduce the likelihood of breaking up a single distribution.

During the PM peak period from 3:00PM – 7:00PM three distributions were fit for the 3:00PM – 5:00PM and 5:00PM – 6:30PM time periods while only two distributions were fit for the 6:30PM – 7:00PM time period. Three distributions for the 3:00PM – 5:00PM time period do a good job describing the data. However, during the 5:00PM – 6:30PM time period three distributions were fit, however, it appears there may be considerable breaking up of what appears to be two underlying distributions. During this period, the AIC value for a three distribution fit was 992 while the next lowest AIC value was 998 for a two distribution fit. Although two underlying distributions exist, it may be possibly that it makes more logical sense to model this as a single distribution with a second distribution capturing the outlying data. During the last time period from 6:30PM – 7:00PM, two distribution were fit and this appears to do a good job describing the data.

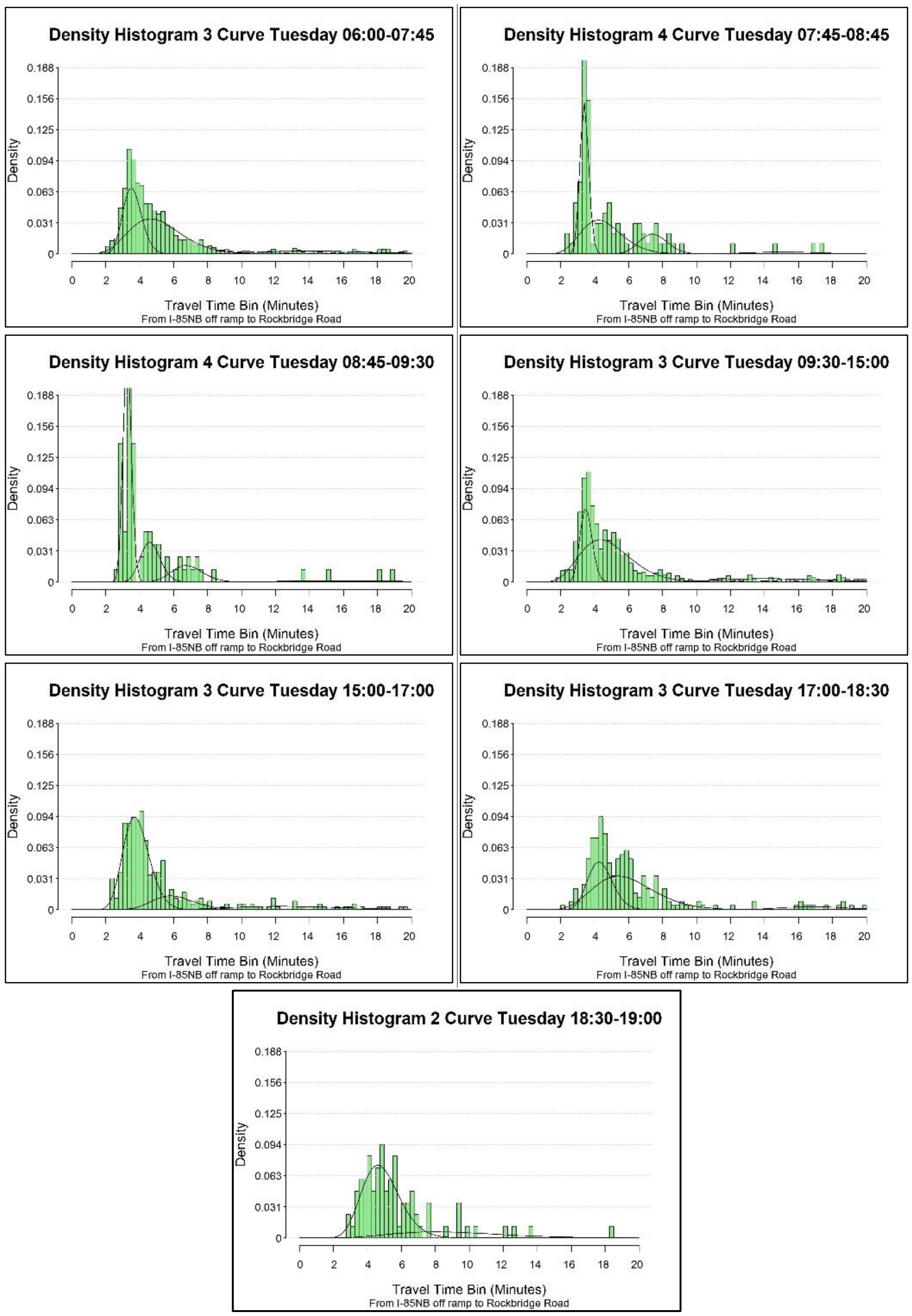


Figure 23: Multi-gamma distribution fit histograms for I-85NB off ramp to Rockbridge Road

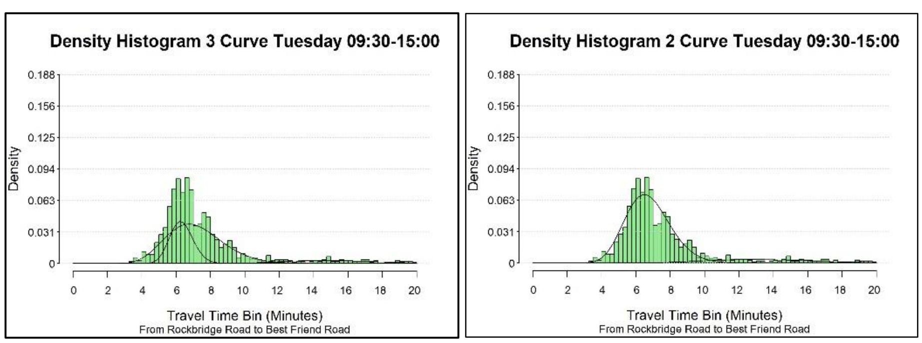
3.3 Distribution Split Spot Check of Overlapping Distributions

As mentioned in a few of the previous sections, there were some instances of the algorithm fitting multiple distributions to the same what visually appears as a single distribution, or distributions that overlap enough where it will break an apparent single distribution into two distributions. This is not ideal as in a before and after analysis one would want to be able to run statistical tests on those individual distributions to determine if any shifts in travel time of the distributions are statistically significant. Furthermore, one would not want to run the risk of finding that an additional distribution was added to the mix or removed from the mix in the after part of a before and after study when this actually did not occur and was just an anomaly of the algorithm.

At its current state, the algorithm will give an optimum solution for one to six gamma distribution fits. This is helpful in determining an alternative solution to a best fit that do not meet the needs of the project. Here, if in a case three distributions breaks a single distribution into two distributions, then one can review the plots for the other fits, as well as the R^2 and AIC value outputs, and decide on the next best fit that may better meet the needs of the project.

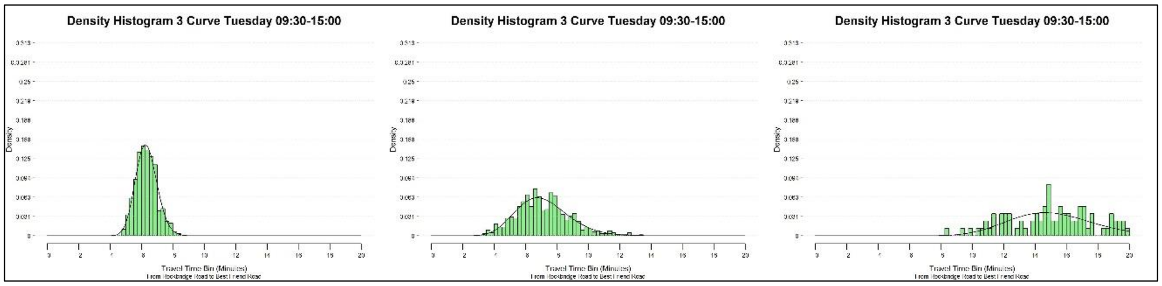
Below in Figure 24a one can see the algorithm's absolute optimum fit, using three distributions for the Rockbridge Road to Best Friend Road route. This solution was the highest R^2 value for the 100 individual three distribution fit and was also the lowest AIC value fit for comparing between one to six distributions. It can be seen that statistically this was the best fit, however, it has two overlapping distributions that will in fact split the first peak into two distributions. This split can be witnessed in Figure 25a where each of the three distributions, and the points assigned to those distributions have been broken apart into separate plots. The first peak at approximately six minutes becomes a good fit, however, the second distribution overlaps the first and consequently removes data points from this peak to use in the second distribution fit shown in Figure 25a in the middle image while the final distribution carries what are most likely outlying data points. One can observe this behavior further by checking the raw travel time data plots, with points colored by distribution, and seeing the overlap of orange and blue points in Figure 26a.

Again, for a before and after analysis this may not be ideal. In order to find what may be a more logical solution, here, one may choose to use only two distributions instead. For the absolute optimum three distributions the AIC value was 4181 while the next lowest AIC value was in fact two distributions at 4201. There is a tradeoff, however, two distributions will group what appears to be a possible second distribution centered at about 8 minutes in with the first distribution centered at approximately 6 minutes. The second distribution is not as prominent as some of the other data for this corridor and it may be better to handle this as a single distribution than to use a best fit that breaks up obvious single distributions. Furthermore, one can see in Figure 25b and in Figure 26b that this solution minimizes breakup of single distributions and allows for separation of possible outlying data. Finally, in Figure 26b one can see a slight rise in travel time across this time period (5.5 hours) which may be the cause of the possible second peak in the density histograms. Another possible approach to this issue may be to identify when this increase in travel times occur by observing the raw data plots and create an additional bin separating this time period into two time periods (from 9:30AM – 12:00PM and 12:00PM – 3:00PM).

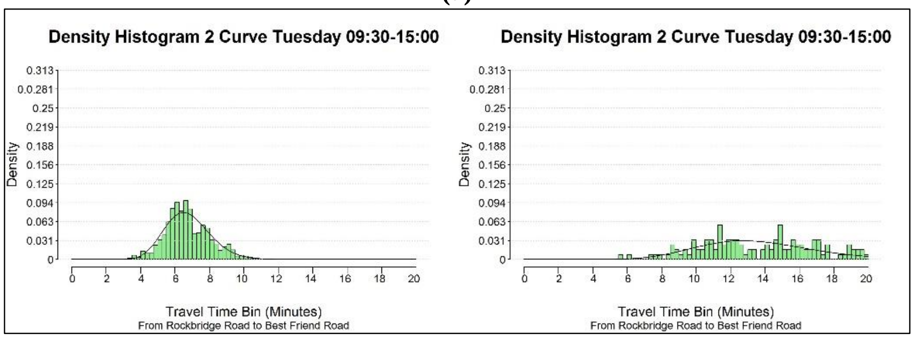


(a) **(b)**

Figure 24: Density histograms for (a) the absolute optimum 3 distribution fit and (b) an optimum 2 distribution fit for the Rockbridge to Best Friend Road route

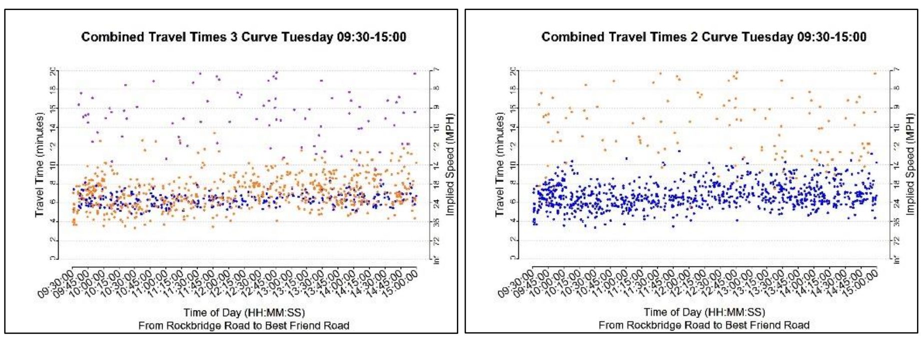


(a)



(b)

Figure 25: Histograms of the separated data by distribution for (a) a 3 distribution fit and (b) a 2 distribution fit for the Rockbridge Road to Best Friend Road route



(a) **(b)**

Figure 26: Travel time raw data plots with data points colored by distribution for the Rockbridge Road to Best Friend Road route for (a) three distributions and (b) two distributions

CHAPTER 4

CONCLUSIONS AND FUTURE RESEARCH

4.1 Conclusions

This research focused on finding a method to fit multiple distributions to travel time data and find the number of underlying distributions in any give set of data. This was of interest as in congested conditions along a signalized corridor or freeway segments, travel time data can often contain multiple underlying distributions. Developing a statistical method to separate these distributions can provide traffic engineers with better tools to more accurately analyze data for various projects. In the data set obtained from Jimmy Carter Boulevard in northeast metropolitan Atlanta, multiple distributions were observed along some routes. These data were used to test the methodology, and may be used to more accurately test differences in travel time data before and after the construction of a new diverging diamond interchange along the corridor. It was desired to use a method to fit multiple gamma distributions, as the gamma distribution is the assumed distribution of travel time data, to the travel time data in order to assign data points to each distribution and separate the data in that manner.

It was not desired to develop our own EM function for fitting multiple distributions for this research as this is considered a mathematical field of study in its own right. Instead, it was desired to utilize an existing tool for a unique problem commonly encountered in the field of traffic engineering. Therefore, an existing tool was found in R's CRAN library repository. The "mixtools" library, from CRAN, was discovered to have a multiple gamma distribution fitting function called the "gammamixEM" function [13]. This function proved to be useful, however, it was discovered that if prior information is not known about the data in order to fit the multiple distributions the function will not necessarily immediately find the global optimum solution. It was discovered after looking at multiple runs of the function that it appeared in most instances to find at least an approximation of the global optimum after many runs although this was determined through visual expectation and is not proven within this document. Therefore, it was chosen to fit each data set 100 times, taking the outputs from the highest R^2 value of each of the 100 runs.

After the highest R^2 results for each number of distributions was found, the best fitting number of distributions was determined using the Akaike's Information Criterion (AIC) to compare between the different number of distributional fits. This statistic uses the log-likelihood of the fit with an adjustment taking into account the number of degrees of freedom, or the number of variables used to fit the distributions, in order to make a comparison between different number of distributional fits. The AIC value allows one to make comparisons on the information content between different fits but does not give information on how good an individual fit is. In our case we assume each case is a best fit by maximizing the R^2 for each number of distributions fit.

For the most part the algorithm developed for this research performed well. Out of 70 overall multiple gamma distribution fits only 10 multiple distribution fits are considered less than ideal due to either overlapping distributions breaking up obvious single distributions or multiple distributions fit to the same peak/distribution breaking up obvious single distributions. However, alternative solutions are available through the process described in section 3.3 on page 48. In this process one can review the outputs of the highest R^2 and lowest AIC values for each number of distributions fit, as well as review their plotted outputs and decide on an alternative solution that best fit the needs of the project.

4.2 Future Research

Future research in this area should focus on the cases where single distribution data splitting by overlapping or multiple distribution fits to a single peak are occurring. Statistically, these are in fact the optimal solutions found that do describe the data well, however, they may not meet the exact needs of some projects where it is not desirable to split up obvious single distributions. The solution to this problem outlined in this research does require some engineering judgement, intentionally choosing a solution that is not the optimal (i.e. lowest R^2) found, and is not a uniform solution across the board. A more desirable solution would require that the statistical output match the logically expected outcome.

Further testing or advancements of the algorithm may involve trying the alternatives specified below:

1. Increase the maximum number of iterations for the fitting function
2. Decrease the convergence criterion of the fitting function
3. Try other distributional types supported by the "mixtools" library with the gamma distribution and compare AIC values across optimized solutions from different distribution types and the number of distributions fit [13].
    a. Check if a different distribution is consistently reporting lower AIC values than the other distribution types
    b. Normal, gamma, logistic, etc.
4. Try alternative statistics for comparing between maximized 100 run outputs
    a. AICc is an AIC correction for finite sample sizes [14]
        i. $AICc = AIC + \frac{2*k*(k+1)}{n-k-1}$

A sensitivity analysis should be performed testing what improvements can be gained by increasing the maximum number of iterations, decreasing the convergence criterion, or both. This could be done using the R^2 outputs from the 100 runs and checking if there is any change in the average R^2 output between runs accepting default values for the maximum number of iterations and convergence criterion and then increasing and decreasing these default values respectively in an attempt to improve the fits. If there is significant improvement in the average R^2 values by changing one or both of these default values, then future use/iterations of the algorithm may want to incorporate these changes.

Alternative distribution types could be checked to see if the final AIC values are improved by using a different distribution type. Using the same "mixtools" library, in which the "gammamixEM" function was found, it appears that one can also choose to use a normal distribution ("normalmixEM") and a logistic distribution ("logisregmixEM") [14]. It could be possible that multiple normal distributions may describe the data better while also benefitting in a reduction in the number of parameters used to fit the functions possibly allowing the AIC values to be lower than a gamma function's.

Finally, future research may want to consider using an AICc instead of a regular AIC. The AICc equation above provides extra penalties for added parameters while also converging toward the regular AIC value when the sample size become large [15]. This statistic was created in order to reduce the likelihood of overfitting a dataset using the AIC values when n is small [15]. It is recommended by Burnham and Anderson to use the AICc value over the AIC value in all cases since it converges toward the regular AIC value when n becomes large. [15].

APPENDIX A: SUMMARY TABLES OF FINAL DISTRIBUTION SELECTIONS

Table A-2: Summary of R^2 and AIC values for final distributions selected for the Best Friend Road to Rockbridge Road route

Day	Number of Distributions	Time Period	R-square	AIC
Tuesday	5	6:00AM - 7:45AM	0.934	14683
Tuesday	4	7:45AM - 8:45AM	0.856	958
Tuesday	5	8:45AM - 9:30AM	0.942	785
Tuesday	5	9:30AM - 3:00PM	0.928	4521
Tuesday	6	3:00PM - 5:00PM	0.944	2544
Tuesday	6	5:00PM - 6:30PM	0.963	2188
Tuesday	5	6:30PM - 7:00PM	0.951	729

Table A-3: Summary of R^2 and AIC values for final distributions selected for the Rockbridge Road to Best Friend Road route

Day	Number of Distributions	Time Period	R-square	AIC
Tuesday	2	6:00AM - 7:45AM	0.536	15136
Tuesday	1	7:45AM - 8:45AM	0.000	2221
Tuesday	2	8:45AM - 9:30AM	0.548	1120
Tuesday	3	9:30AM - 3:00PM	0.694	4181
Tuesday	1	3:00PM - 5:00PM	0.000	1579
Tuesday	2	5:00PM - 6:30PM	0.551	1081
Tuesday	2	6:30PM - 7:00PM	0.675	399

Table A-4: Summary of R^2 and AIC values for final distributions selected for the Best Friend Road to I-85SB on ramp route

Day	Number of Distributions	Time Period	R-square	AIC
Tuesday	6	6:00AM - 7:45AM	0.928	15373
Tuesday	2	7:45AM - 8:45AM	0.659	1314
Tuesday	3	8:45AM - 9:30AM	0.745	1058
Tuesday	6	9:30AM - 3:00PM	0.952	6235
Tuesday	4	3:00PM - 5:00PM	0.896	2099
Tuesday	1	5:00PM - 6:30PM	0.000	1581
Tuesday	3	6:30PM - 7:00PM	0.896	430

Table A-5: Summary of R^2 and AIC values for final distributions selected for the Best Friend Road to I-85NB on ramp route

Day	Number of Distributions	Time Period	R-square	AIC
Tuesday	5	6:00AM - 7:45AM	0.907	22249
Tuesday	5	7:45AM - 8:45AM	0.933	1782
Tuesday	2	8:45AM - 9:30AM	0.393	1260
Tuesday	5	9:30AM - 3:00PM	0.916	8148
Tuesday	2	3:00PM - 5:00PM	0.569	3174
Tuesday	1	5:00PM - 6:30PM	0.000	2489
Tuesday	5	6:30PM - 7:00PM	0.940	768

Table A-6: Summary of R^2 and AIC values for final distributions selected for the Rockbridge Road to I-85SB on ramp route

Day	Number of Distributions	Time Period	R-square	AIC
Tuesday	3	6:00AM - 7:45AM	0.794	11948
Tuesday	2	7:45AM - 8:45AM	0.649	1543
Tuesday	2	8:45AM - 9:30AM	0.725	898
Tuesday	3	9:30AM - 3:00PM	0.832	4643
Tuesday	3	3:00PM - 5:00PM	0.731	1138
Tuesday	2	5:00PM - 6:30PM	0.710	682
Tuesday	2	6:30PM - 7:00PM	0.741	274

Table A-7: Summary of R^2 and AIC values for final distributions selected for the Rockbridge Road to I-85NB on ramp route

Day	Number of Distributions	Time Period	R-square	AIC
Tuesday	4	6:00AM - 7:45AM	0.785	16234
Tuesday	5	7:45AM - 8:45AM	0.891	2015
Tuesday	2	8:45AM - 9:30AM	0.579	1082
Tuesday	5	9:30AM - 3:00PM	0.905	5857
Tuesday	6	3:00PM - 5:00PM	0.951	2215
Tuesday	4	5:00PM - 6:30PM	0.847	1395
Tuesday	4	6:30PM - 7:00PM	0.879	440

Table A-8: Summary of R^2 and AIC values for final distributions selected for the I-85SB off ramp to Best Friend Road route

Day	Number of Distributions	Time Period	R-square	AIC
Tuesday	3	6:00AM - 7:45AM	0.849	18859
Tuesday	3	7:45AM - 8:45AM	0.896	1474
Tuesday	4	8:45AM - 9:30AM	0.899	1066
Tuesday	4	9:30AM - 3:00PM	0.884	8420
Tuesday	2	3:00PM - 5:00PM	0.745	2354
Tuesday	3	5:00PM - 6:30PM	0.754	1483
Tuesday	3	6:30PM - 7:00PM	0.869	516

Table A-9: Summary of R^2 and AIC values for final distributions selected for the I-85 SB off ramp to Rockbridge Road route

Day	Number of Distributions	Time Period	R-square	AIC
Tuesday	3	6:00AM - 7:45AM	0.755	10156
Tuesday	3	7:45AM - 8:45AM	0.886	413
Tuesday	3	8:45AM - 9:30AM	0.862	383
Tuesday	3	9:30AM - 3:00PM	0.827	4189
Tuesday	3	3:00PM - 5:00PM	0.791	1791
Tuesday	5	5:00PM - 6:30PM	0.941	1406
Tuesday	4	6:30PM - 7:00PM	0.921	438

Table A-10: Summary of R^2 and AIC values for final distributions selected for the I-85NB off ramp to Best Friend Road route

Day	Number of Distributions	Time Period	R-square	AIC
Tuesday	3	6:00AM - 7:45AM	0.865	7532
Tuesday	3	7:45AM - 8:45AM	0.900	582
Tuesday	3	8:45AM - 9:30AM	0.903	567
Tuesday	3	9:30AM - 3:00PM	0.893	3692
Tuesday	2	3:00PM - 5:00PM	0.720	950
Tuesday	2	5:00PM - 6:30PM	0.805	468
Tuesday	2	6:30PM - 7:00PM	0.851	156

Table A-11: Summary of R^2 and AIC values for final distributions selected for the I-85NB off ramp to Rockbridge Road route

Day	Number of Distributions	Time Period	R-square	AIC
Tuesday	3	6:00AM - 7:45AM	0.803	7639
Tuesday	4	7:45AM - 8:45AM	0.883	367
Tuesday	4	8:45AM - 9:30AM	0.956	279
Tuesday	3	9:30AM - 3:00PM	0.810	3250
Tuesday	3	3:00PM - 5:00PM	0.844	1368
Tuesday	3	5:00PM - 6:30PM	0.781	992
Tuesday	2	6:30PM - 7:00PM	0.621	354

APPENDIX B: TABLES OF MAXIMIZED R^2 AND AIC VALUES

Table B-12: Table of R^2 and AIC values for maximized R^2 output of each distribution for the Best Friend Road to Rockbridge Road route

Day	Number of Distributions	Hour	R-square	AIC
Tuesday	1	6:00AM - 7:45AM	0.000	15397
Tuesday	2	6:00AM - 7:45AM	0.502	14880
Tuesday	3	6:00AM - 7:45AM	0.756	14880
Tuesday	4	6:00AM - 7:45AM	0.887	14752
Tuesday	5	6:00AM - 7:45AM	0.934	14683
Tuesday	6	6:00AM - 7:45AM	0.939	14925
Tuesday	1	7:45AM - 8:45AM	0.000	997
Tuesday	2	7:45AM - 8:45AM	0.455	989
Tuesday	3	7:45AM - 8:45AM	0.769	991
Tuesday	4	7:45AM - 8:45AM	0.856	958
Tuesday	5	7:45AM - 8:45AM	0.934	985
Tuesday	6	7:45AM - 8:45AM	0.950	975
Tuesday	1	8:45AM - 9:30AM	0.000	800
Tuesday	2	8:45AM - 9:30AM	0.464	804
Tuesday	3	8:45AM - 9:30AM	0.742	809
Tuesday	4	8:45AM - 9:30AM	0.879	817
Tuesday	5	8:45AM - 9:30AM	0.942	785
Tuesday	6	8:45AM - 9:30AM	0.942	798
Tuesday	1	9:30AM - 3:00PM	0.000	4979
Tuesday	2	9:30AM - 3:00PM	0.677	4639
Tuesday	3	9:30AM - 3:00PM	0.782	4616
Tuesday	4	9:30AM - 3:00PM	0.894	4557
Tuesday	5	9:30AM - 3:00PM	0.928	4521
Tuesday	6	9:30AM - 3:00PM	0.930	4607
Tuesday	1	3:00PM - 5:00PM	0.000	2804
Tuesday	2	3:00PM - 5:00PM	0.437	2720
Tuesday	3	3:00PM - 5:00PM	0.702	2600
Tuesday	4	3:00PM - 5:00PM	0.892	2636
Tuesday	5	3:00PM - 5:00PM	0.909	2621
Tuesday	6	3:00PM - 5:00PM	0.944	2544
Tuesday	1	5:00PM - 6:30PM	0.000	2294
Tuesday	2	5:00PM - 6:30PM	0.473	2298
Tuesday	3	5:00PM - 6:30PM	0.790	2318
Tuesday	4	5:00PM - 6:30PM	0.911	2244
Tuesday	5	5:00PM - 6:30PM	0.936	2247
Tuesday	6	5:00PM - 6:30PM	0.963	2188
Tuesday	1	6:30PM - 7:00PM	0.000	753
Tuesday	2	6:30PM - 7:00PM	0.592	743
Tuesday	3	6:30PM - 7:00PM	0.843	752
Tuesday	4	6:30PM - 7:00PM	0.909	756
Tuesday	5	6:30PM - 7:00PM	0.951	729
Tuesday	6	6:30PM - 7:00PM	0.963	741

Table B-13: Table of R^2 and AIC values for maximized R^2 output of each distribution for the Rockbridge Road to Best Friend Route

Day	Distribution Number	Hour	R-square	AIC
Tuesday	1	6:00AM - 7:45AM	0.000	15365
Tuesday	2	6:00AM - 7:45AM	0.536	15136
Tuesday	3	6:00AM - 7:45AM	0.651	15209
Tuesday	4	6:00AM - 7:45AM	0.805	15380
Tuesday	5	6:00AM - 7:45AM	0.888	15283
Tuesday	6	6:00AM - 7:45AM	0.919	15394
Tuesday	1	7:45AM - 8:45AM	0.000	2221
Tuesday	2	7:45AM - 8:45AM	0.295	2228
Tuesday	3	7:45AM - 8:45AM	0.722	2227
Tuesday	4	7:45AM - 8:45AM	0.859	2259
Tuesday	5	7:45AM - 8:45AM	0.917	2247
Tuesday	6	7:45AM - 8:45AM	0.940	2263
Tuesday	1	8:45AM - 9:30AM	0.000	1178
Tuesday	2	8:45AM - 9:30AM	0.548	1120
Tuesday	3	8:45AM - 9:30AM	0.699	1126
Tuesday	4	8:45AM - 9:30AM	0.797	1149
Tuesday	5	8:45AM - 9:30AM	0.889	1152
Tuesday	6	8:45AM - 9:30AM	0.899	1139
Tuesday	1	9:30AM - 3:00PM	0.000	4576
Tuesday	2	9:30AM - 3:00PM	0.672	4201
Tuesday	3	9:30AM - 3:00PM	0.694	4181
Tuesday	4	9:30AM - 3:00PM	0.867	4306
Tuesday	5	9:30AM - 3:00PM	0.879	4346
Tuesday	6	9:30AM - 3:00PM	0.923	4284
Tuesday	1	3:00PM - 5:00PM	0.000	1579
Tuesday	2	3:00PM - 5:00PM	0.450	1583
Tuesday	3	3:00PM - 5:00PM	0.683	1599
Tuesday	4	3:00PM - 5:00PM	0.854	1608
Tuesday	5	3:00PM - 5:00PM	0.925	1596
Tuesday	6	3:00PM - 5:00PM	0.947	1584
Tuesday	1	5:00PM - 6:30PM	0.000	1116
Tuesday	2	5:00PM - 6:30PM	0.551	1081
Tuesday	3	5:00PM - 6:30PM	0.684	1088
Tuesday	4	5:00PM - 6:30PM	0.897	1096
Tuesday	5	5:00PM - 6:30PM	0.927	1103
Tuesday	6	5:00PM - 6:30PM	0.951	1102
Tuesday	1	6:30PM - 7:00PM	0.000	409
Tuesday	2	6:30PM - 7:00PM	0.675	399
Tuesday	3	6:30PM - 7:00PM	0.762	404
Tuesday	4	6:30PM - 7:00PM	0.867	410
Tuesday	5	6:30PM - 7:00PM	0.925	415
Tuesday	6	6:30PM - 7:00PM	0.948	421

Table B-14: Table of R^2 and AIC values for maximized R^2 output of each distribution for the Best Friend to I-85SB on ramp route

Day	Distribution Number	Hour	R-square	chi-sqr
Tuesday	1	6:00AM - 7:45AM	0.000	23105
Tuesday	2	6:00AM - 7:45AM	0.419	22554
Tuesday	3	6:00AM - 7:45AM	0.736	22520
Tuesday	4	6:00AM - 7:45AM	0.814	22856
Tuesday	5	6:00AM - 7:45AM	0.907	22249
Tuesday	6	6:00AM - 7:45AM	0.949	22446
Tuesday	1	7:45AM - 8:45AM	0.000	1799
Tuesday	2	7:45AM - 8:45AM	0.363	1796
Tuesday	3	7:45AM - 8:45AM	0.793	1788
Tuesday	4	7:45AM - 8:45AM	0.900	1789
Tuesday	5	7:45AM - 8:45AM	0.933	1782
Tuesday	6	7:45AM - 8:45AM	0.960	1784
Tuesday	1	8:45AM - 9:30AM	0.000	1265
Tuesday	2	8:45AM - 9:30AM	0.393	1260
Tuesday	3	8:45AM - 9:30AM	0.671	1265
Tuesday	4	8:45AM - 9:30AM	0.833	1274
Tuesday	5	8:45AM - 9:30AM	0.893	1262
Tuesday	6	8:45AM - 9:30AM	0.918	1276
Tuesday	1	9:30AM - 3:00PM	0.000	9268
Tuesday	2	9:30AM - 3:00PM	0.490	8688
Tuesday	3	9:30AM - 3:00PM	0.722	8656
Tuesday	4	9:30AM - 3:00PM	0.858	8193
Tuesday	5	9:30AM - 3:00PM	0.916	8148
Tuesday	6	9:30AM - 3:00PM	0.915	8721
Tuesday	1	3:00PM - 5:00PM	0.000	3261
Tuesday	2	3:00PM - 5:00PM	0.569	3174
Tuesday	3	3:00PM - 5:00PM	0.638	3175
Tuesday	4	3:00PM - 5:00PM	0.798	3262
Tuesday	5	3:00PM - 5:00PM	0.870	3183
Tuesday	6	3:00PM - 5:00PM	0.909	3249
Tuesday	1	5:00PM - 6:30PM	0.000	2489
Tuesday	2	5:00PM - 6:30PM	0.388	2493
Tuesday	3	5:00PM - 6:30PM	0.669	2553
Tuesday	4	5:00PM - 6:30PM	0.821	2528
Tuesday	5	5:00PM - 6:30PM	0.893	2559
Tuesday	6	5:00PM - 6:30PM	0.927	2562
Tuesday	1	6:30PM - 7:00PM	0.000	775
Tuesday	2	6:30PM - 7:00PM	0.704	769
Tuesday	3	6:30PM - 7:00PM	0.772	768
Tuesday	4	6:30PM - 7:00PM	0.905	773
Tuesday	5	6:30PM - 7:00PM	0.940	768
Tuesday	6	6:30PM - 7:00PM	0.962	778

Table B-15: Table of R^2 and AIC values for maximized R^2 output of each distribution for the Best Friend Road to I-85NB on ramp route

Day	Distribution Number	Hour	R-square	chi-sqr
Tuesday	1	6:00AM - 7:45AM	0.000	16553
Tuesday	2	6:00AM - 7:45AM	0.433	15614
Tuesday	3	6:00AM - 7:45AM	0.759	15438
Tuesday	4	6:00AM - 7:45AM	0.813	15461
Tuesday	5	6:00AM - 7:45AM	0.906	15414
Tuesday	6	6:00AM - 7:45AM	0.928	15373
Tuesday	1	7:45AM - 8:45AM	0.000	1360
Tuesday	2	7:45AM - 8:45AM	0.659	1314
Tuesday	3	7:45AM - 8:45AM	0.703	1320
Tuesday	4	7:45AM - 8:45AM	0.860	1316
Tuesday	5	7:45AM - 8:45AM	0.907	1354
Tuesday	6	7:45AM - 8:45AM	0.923	1326
Tuesday	1	8:45AM - 9:30AM	0.000	1076
Tuesday	2	8:45AM - 9:30AM	0.632	1061
Tuesday	3	8:45AM - 9:30AM	0.745	1058
Tuesday	4	8:45AM - 9:30AM	0.799	1060
Tuesday	5	8:45AM - 9:30AM	0.897	1105
Tuesday	6	8:45AM - 9:30AM	0.926	1066
Tuesday	1	9:30AM - 3:00PM	0.000	7330
Tuesday	2	9:30AM - 3:00PM	0.542	6339
Tuesday	3	9:30AM - 3:00PM	0.835	6277
Tuesday	4	9:30AM - 3:00PM	0.903	6267
Tuesday	5	9:30AM - 3:00PM	0.910	6262
Tuesday	6	9:30AM - 3:00PM	0.952	6235
Tuesday	1	3:00PM - 5:00PM	0.000	2467
Tuesday	2	3:00PM - 5:00PM	0.737	2130
Tuesday	3	3:00PM - 5:00PM	0.853	2120
Tuesday	4	3:00PM - 5:00PM	0.896	2099
Tuesday	5	3:00PM - 5:00PM	0.919	2127
Tuesday	6	3:00PM - 5:00PM	0.941	2122
Tuesday	1	5:00PM - 6:30PM	0.000	1581
Tuesday	2	5:00PM - 6:30PM	0.358	1584
Tuesday	3	5:00PM - 6:30PM	0.609	1619
Tuesday	4	5:00PM - 6:30PM	0.806	1628
Tuesday	5	5:00PM - 6:30PM	0.882	1654
Tuesday	6	5:00PM - 6:30PM	0.918	1626
Tuesday	1	6:30PM - 7:00PM	0.000	447
Tuesday	2	6:30PM - 7:00PM	0.731	434
Tuesday	3	6:30PM - 7:00PM	0.896	430
Tuesday	4	6:30PM - 7:00PM	0.939	434
Tuesday	5	6:30PM - 7:00PM	0.952	438
Tuesday	6	6:30PM - 7:00PM	0.952	448

Table B-16: Table of R^2 and AIC values for maximized R^2 output of each distribution for the Rockbridge Road to I-85SB route

Day	Distribution Number	Hour	R-square	chi-sqr
Tuesday	1	6:00AM - 7:45AM	0.000	12868
Tuesday	2	6:00AM - 7:45AM	0.619	12008
Tuesday	3	6:00AM - 7:45AM	0.794	11948
Tuesday	4	6:00AM - 7:45AM	0.809	11965
Tuesday	5	6:00AM - 7:45AM	0.910	12091
Tuesday	6	6:00AM - 7:45AM	0.937	12228
Tuesday	1	7:45AM - 8:45AM	0.000	1588
Tuesday	2	7:45AM - 8:45AM	0.649	1543
Tuesday	3	7:45AM - 8:45AM	0.798	1549
Tuesday	4	7:45AM - 8:45AM	0.880	1573
Tuesday	5	7:45AM - 8:45AM	0.944	1579
Tuesday	6	7:45AM - 8:45AM	0.942	1599
Tuesday	1	8:45AM - 9:30AM	0.000	953
Tuesday	2	8:45AM - 9:30AM	0.725	898
Tuesday	3	8:45AM - 9:30AM	0.773	904
Tuesday	4	8:45AM - 9:30AM	0.854	929
Tuesday	5	8:45AM - 9:30AM	0.909	946
Tuesday	6	8:45AM - 9:30AM	0.947	927
Tuesday	1	9:30AM - 3:00PM	0.000	5317
Tuesday	2	9:30AM - 3:00PM	0.754	4662
Tuesday	3	9:30AM - 3:00PM	0.832	4643
Tuesday	4	9:30AM - 3:00PM	0.826	4893
Tuesday	5	9:30AM - 3:00PM	0.876	4779
Tuesday	6	9:30AM - 3:00PM	0.948	4712
Tuesday	1	3:00PM - 5:00PM	0.000	1265
Tuesday	2	3:00PM - 5:00PM	0.624	1192
Tuesday	3	3:00PM - 5:00PM	0.731	1138
Tuesday	4	3:00PM - 5:00PM	0.855	1156
Tuesday	5	3:00PM - 5:00PM	0.922	1185
Tuesday	6	3:00PM - 5:00PM	0.944	1212
Tuesday	1	5:00PM - 6:30PM	0.000	750
Tuesday	2	5:00PM - 6:30PM	0.710	682
Tuesday	3	5:00PM - 6:30PM	0.833	684
Tuesday	4	5:00PM - 6:30PM	0.905	691
Tuesday	5	5:00PM - 6:30PM	0.919	701
Tuesday	6	5:00PM - 6:30PM	0.941	715
Tuesday	1	6:30PM - 7:00PM	0.000	293
Tuesday	2	6:30PM - 7:00PM	0.741	274
Tuesday	3	6:30PM - 7:00PM	0.908	276
Tuesday	4	6:30PM - 7:00PM	0.930	281
Tuesday	5	6:30PM - 7:00PM	0.950	293
Tuesday	6	6:30PM - 7:00PM	0.947	303

Table B-17: Table of R^2 and AIC values for maximized R^2 output of each distribution for the Rockbridge Road to I-85NB route

Day	Distribution Number	Hour	R-square	chi-sqr
Tuesday	1	6:00AM - 7:45AM	0.000	17564
Tuesday	2	6:00AM - 7:45AM	0.472	16556
Tuesday	3	6:00AM - 7:45AM	0.767	16514
Tuesday	4	6:00AM - 7:45AM	0.785	16234
Tuesday	5	6:00AM - 7:45AM	0.881	16586
Tuesday	6	6:00AM - 7:45AM	0.893	16538
Tuesday	1	7:45AM - 8:45AM	0.000	2064
Tuesday	2	7:45AM - 8:45AM	0.379	2058
Tuesday	3	7:45AM - 8:45AM	0.776	2054
Tuesday	4	7:45AM - 8:45AM	0.830	2048
Tuesday	5	7:45AM - 8:45AM	0.891	2015
Tuesday	6	7:45AM - 8:45AM	0.918	2024
Tuesday	1	8:45AM - 9:30AM	0.000	1155
Tuesday	2	8:45AM - 9:30AM	0.579	1082
Tuesday	3	8:45AM - 9:30AM	0.748	1086
Tuesday	4	8:45AM - 9:30AM	0.780	1090
Tuesday	5	8:45AM - 9:30AM	0.842	1090
Tuesday	6	8:45AM - 9:30AM	0.908	1100
Tuesday	1	9:30AM - 3:00PM	0.000	6917
Tuesday	2	9:30AM - 3:00PM	0.711	5910
Tuesday	3	9:30AM - 3:00PM	0.804	5891
Tuesday	4	9:30AM - 3:00PM	0.807	6040
Tuesday	5	9:30AM - 3:00PM	0.905	5857
Tuesday	6	9:30AM - 3:00PM	0.914	6024
Tuesday	1	3:00PM - 5:00PM	0.000	2378
Tuesday	2	3:00PM - 5:00PM	0.559	2249
Tuesday	3	3:00PM - 5:00PM	0.758	2240
Tuesday	4	3:00PM - 5:00PM	0.810	2237
Tuesday	5	3:00PM - 5:00PM	0.853	2259
Tuesday	6	3:00PM - 5:00PM	0.951	2215
Tuesday	1	5:00PM - 6:30PM	0.000	1430
Tuesday	2	5:00PM - 6:30PM	0.598	1402
Tuesday	3	5:00PM - 6:30PM	0.685	1410
Tuesday	4	5:00PM - 6:30PM	0.847	1395
Tuesday	5	5:00PM - 6:30PM	0.910	1400
Tuesday	6	5:00PM - 6:30PM	0.934	1419
Tuesday	1	6:30PM - 7:00PM	0.000	451
Tuesday	2	6:30PM - 7:00PM	0.524	446
Tuesday	3	6:30PM - 7:00PM	0.653	452
Tuesday	4	6:30PM - 7:00PM	0.879	440
Tuesday	5	6:30PM - 7:00PM	0.899	445
Tuesday	6	6:30PM - 7:00PM	0.948	445

Table B-18: Table of R^2 and AIC values for maximized R^2 output of each distribution for the I-85SB off ramp to Best Friend Road route

Day	Distribution Number	Hour	R-square	chi-sqr
Tuesday	1	6:00AM - 7:45AM	0.000	23028
Tuesday	2	6:00AM - 7:45AM	0.770	18952
Tuesday	3	6:00AM - 7:45AM	0.849	18859
Tuesday	4	6:00AM - 7:45AM	0.881	19616
Tuesday	5	6:00AM - 7:45AM	0.916	19129
Tuesday	6	6:00AM - 7:45AM	0.944	19552
Tuesday	1	7:45AM - 8:45AM	0.000	1912
Tuesday	2	7:45AM - 8:45AM	0.793	1476
Tuesday	3	7:45AM - 8:45AM	0.896	1474
Tuesday	4	7:45AM - 8:45AM	0.889	1507
Tuesday	5	7:45AM - 8:45AM	0.925	1546
Tuesday	6	7:45AM - 8:45AM	0.934	1552
Tuesday	1	8:45AM - 9:30AM	0.000	1396
Tuesday	2	8:45AM - 9:30AM	0.789	1094
Tuesday	3	8:45AM - 9:30AM	0.905	1092
Tuesday	4	8:45AM - 9:30AM	0.899	1066
Tuesday	5	8:45AM - 9:30AM	0.936	1071
Tuesday	6	8:45AM - 9:30AM	0.932	1105
Tuesday	1	9:30AM - 3:00PM	0.000	10591
Tuesday	2	9:30AM - 3:00PM	0.768	8484
Tuesday	3	9:30AM - 3:00PM	0.853	8433
Tuesday	4	9:30AM - 3:00PM	0.884	8420
Tuesday	5	9:30AM - 3:00PM	0.881	8801
Tuesday	6	9:30AM - 3:00PM	0.935	8850
Tuesday	1	3:00PM - 5:00PM	0.000	2891
Tuesday	2	3:00PM - 5:00PM	0.745	2354
Tuesday	3	3:00PM - 5:00PM	0.805	2471
Tuesday	4	3:00PM - 5:00PM	0.851	2504
Tuesday	5	3:00PM - 5:00PM	0.921	2454
Tuesday	6	3:00PM - 5:00PM	0.907	2540
Tuesday	1	5:00PM - 6:30PM	0.000	1902
Tuesday	2	5:00PM - 6:30PM	0.775	1483
Tuesday	3	5:00PM - 6:30PM	0.754	1483
Tuesday	4	5:00PM - 6:30PM	0.859	1633
Tuesday	5	5:00PM - 6:30PM	0.869	1634
Tuesday	6	5:00PM - 6:30PM	0.896	1592
Tuesday	1	6:30PM - 7:00PM	0.000	679
Tuesday	2	6:30PM - 7:00PM	0.789	520
Tuesday	3	6:30PM - 7:00PM	0.869	516
Tuesday	4	6:30PM - 7:00PM	0.888	574
Tuesday	5	6:30PM - 7:00PM	0.897	583
Tuesday	6	6:30PM - 7:00PM	0.928	550

Table B-19: Table of R^2 and AIC values for maximized R^2 output of each distribution for the I-85SB off ramp to Rockbridge Road route

Day	Distribution Number	Hour	R-square	chi-sqr
Tuesday	1	6:00AM - 7:45AM	0.000	11245
Tuesday	2	6:00AM - 7:45AM	0.528	10402
Tuesday	3	6:00AM - 7:45AM	0.755	10156
Tuesday	4	6:00AM - 7:45AM	0.890	10174
Tuesday	5	6:00AM - 7:45AM	0.886	10373
Tuesday	6	6:00AM - 7:45AM	0.932	10388
Tuesday	1	7:45AM - 8:45AM	0.000	517
Tuesday	2	7:45AM - 8:45AM	0.395	426
Tuesday	3	7:45AM - 8:45AM	0.886	413
Tuesday	4	7:45AM - 8:45AM	0.865	418
Tuesday	5	7:45AM - 8:45AM	0.914	422
Tuesday	6	7:45AM - 8:45AM	0.906	432
Tuesday	1	8:45AM - 9:30AM	0.000	476
Tuesday	2	8:45AM - 9:30AM	0.395	404
Tuesday	3	8:45AM - 9:30AM	0.862	383
Tuesday	4	8:45AM - 9:30AM	0.868	395
Tuesday	5	8:45AM - 9:30AM	0.884	405
Tuesday	6	8:45AM - 9:30AM	0.893	414
Tuesday	1	9:30AM - 3:00PM	0.000	4760
Tuesday	2	9:30AM - 3:00PM	0.772	4197
Tuesday	3	9:30AM - 3:00PM	0.827	4189
Tuesday	4	9:30AM - 3:00PM	0.880	4252
Tuesday	5	9:30AM - 3:00PM	0.897	4320
Tuesday	6	9:30AM - 3:00PM	0.923	4311
Tuesday	1	3:00PM - 5:00PM	0.000	1898
Tuesday	2	3:00PM - 5:00PM	0.661	1805
Tuesday	3	3:00PM - 5:00PM	0.791	1791
Tuesday	4	3:00PM - 5:00PM	0.851	1792
Tuesday	5	3:00PM - 5:00PM	0.891	1850
Tuesday	6	3:00PM - 5:00PM	0.901	1859
Tuesday	1	5:00PM - 6:30PM	0.000	1472
Tuesday	2	5:00PM - 6:30PM	0.648	1424
Tuesday	3	5:00PM - 6:30PM	0.728	1423
Tuesday	4	5:00PM - 6:30PM	0.899	1412
Tuesday	5	5:00PM - 6:30PM	0.941	1406
Tuesday	6	5:00PM - 6:30PM	0.951	1429
Tuesday	1	6:30PM - 7:00PM	0.000	453
Tuesday	2	6:30PM - 7:00PM	0.600	448
Tuesday	3	6:30PM - 7:00PM	0.856	452
Tuesday	4	6:30PM - 7:00PM	0.921	438
Tuesday	5	6:30PM - 7:00PM	0.958	439
Tuesday	6	6:30PM - 7:00PM	0.954	446

Table B-20: Table of R^2 and AIC values for maximized R^2 output of each distribution for the I-85NB off ramp to Best Friend Road route

Day	Distribution Number	Hour	R-square	chi-sqr
Tuesday	1	6:00AM - 7:45AM	0.000	8228
Tuesday	2	6:00AM - 7:45AM	0.780	7572
Tuesday	3	6:00AM - 7:45AM	0.865	7532
Tuesday	4	6:00AM - 7:45AM	0.877	7545
Tuesday	5	6:00AM - 7:45AM	0.929	7599
Tuesday	6	6:00AM - 7:45AM	0.949	7588
Tuesday	1	7:45AM - 8:45AM	0.000	641
Tuesday	2	7:45AM - 8:45AM	0.774	584
Tuesday	3	7:45AM - 8:45AM	0.900	582
Tuesday	4	7:45AM - 8:45AM	0.932	584
Tuesday	5	7:45AM - 8:45AM	0.945	590
Tuesday	6	7:45AM - 8:45AM	0.960	611
Tuesday	1	8:45AM - 9:30AM	0.000	616
Tuesday	2	8:45AM - 9:30AM	0.797	567
Tuesday	3	8:45AM - 9:30AM	0.903	567
Tuesday	4	8:45AM - 9:30AM	0.928	570
Tuesday	5	8:45AM - 9:30AM	0.945	571
Tuesday	6	8:45AM - 9:30AM	0.970	584
Tuesday	1	9:30AM - 3:00PM	0.000	4111
Tuesday	2	9:30AM - 3:00PM	0.793	3718
Tuesday	3	9:30AM - 3:00PM	0.893	3692
Tuesday	4	9:30AM - 3:00PM	0.889	3716
Tuesday	5	9:30AM - 3:00PM	0.932	3792
Tuesday	6	9:30AM - 3:00PM	0.950	3797
Tuesday	1	3:00PM - 5:00PM	0.000	975
Tuesday	2	3:00PM - 5:00PM	0.720	950
Tuesday	3	3:00PM - 5:00PM	0.750	958
Tuesday	4	3:00PM - 5:00PM	0.872	961
Tuesday	5	3:00PM - 5:00PM	0.908	982
Tuesday	6	3:00PM - 5:00PM	0.946	990
Tuesday	1	5:00PM - 6:30PM	0.000	510
Tuesday	2	5:00PM - 6:30PM	0.805	468
Tuesday	3	5:00PM - 6:30PM	0.875	473
Tuesday	4	5:00PM - 6:30PM	0.894	480
Tuesday	5	5:00PM - 6:30PM	0.944	485
Tuesday	6	5:00PM - 6:30PM	0.940	495
Tuesday	1	6:30PM - 7:00PM	0.000	171
Tuesday	2	6:30PM - 7:00PM	0.851	156
Tuesday	3	6:30PM - 7:00PM	0.928	160
Tuesday	4	6:30PM - 7:00PM	0.972	169
Tuesday	5	6:30PM - 7:00PM	0.975	174
Tuesday	6	6:30PM - 7:00PM	0.981	174

Table B-21: Table of R^2 and AIC values for maximized R^2 output of each distribution for the I-85NB off ramp to Rockbridge Road route

Day	Distribution Number	Hour	R-square	chi-sqr
Tuesday	1	6:00AM - 7:45AM	0.000	8726
Tuesday	2	6:00AM - 7:45AM	0.555	7763
Tuesday	3	6:00AM - 7:45AM	0.803	7639
Tuesday	4	6:00AM - 7:45AM	0.868	7644
Tuesday	5	6:00AM - 7:45AM	0.925	7833
Tuesday	6	6:00AM - 7:45AM	0.928	7913
Tuesday	1	7:45AM - 8:45AM	0.000	424
Tuesday	2	7:45AM - 8:45AM	0.613	406
Tuesday	3	7:45AM - 8:45AM	0.869	376
Tuesday	4	7:45AM - 8:45AM	0.883	367
Tuesday	5	7:45AM - 8:45AM	0.940	381
Tuesday	6	7:45AM - 8:45AM	0.946	387
Tuesday	1	8:45AM - 9:30AM	0.000	348
Tuesday	2	8:45AM - 9:30AM	0.781	308
Tuesday	3	8:45AM - 9:30AM	0.924	290
Tuesday	4	8:45AM - 9:30AM	0.956	279
Tuesday	5	8:45AM - 9:30AM	0.949	289
Tuesday	6	8:45AM - 9:30AM	0.963	279
Tuesday	1	9:30AM - 3:00PM	0.000	3751
Tuesday	2	9:30AM - 3:00PM	0.639	3309
Tuesday	3	9:30AM - 3:00PM	0.810	3250
Tuesday	4	9:30AM - 3:00PM	0.891	3270
Tuesday	5	9:30AM - 3:00PM	0.925	3391
Tuesday	6	9:30AM - 3:00PM	0.944	3287
Tuesday	1	3:00PM - 5:00PM	0.000	1574
Tuesday	2	3:00PM - 5:00PM	0.616	1370
Tuesday	3	3:00PM - 5:00PM	0.844	1368
Tuesday	4	3:00PM - 5:00PM	0.859	1384
Tuesday	5	3:00PM - 5:00PM	0.913	1417
Tuesday	6	3:00PM - 5:00PM	0.915	1414
Tuesday	1	5:00PM - 6:30PM	0.000	1094
Tuesday	2	5:00PM - 6:30PM	0.737	998
Tuesday	3	5:00PM - 6:30PM	0.781	992
Tuesday	4	5:00PM - 6:30PM	0.911	1003
Tuesday	5	5:00PM - 6:30PM	0.916	1008
Tuesday	6	5:00PM - 6:30PM	0.950	1017
Tuesday	1	6:30PM - 7:00PM	0.000	371
Tuesday	2	6:30PM - 7:00PM	0.621	354
Tuesday	3	6:30PM - 7:00PM	0.681	360
Tuesday	4	6:30PM - 7:00PM	0.862	367
Tuesday	5	6:30PM - 7:00PM	0.873	376
Tuesday	6	6:30PM - 7:00PM	0.915	389

REFERENCES

[1] Ji, Y. Zhang, H. M. Travel Time Distributions on Urban Streets: Their Estimation with a Hierarchical Bayesian Mixture Model and Application to Traffic Analysis Using High-Resolution Bus Probe Data. In *Transportation Research Board 92nd Annual Meeting*. Transportation Research Board of the National Academies. Washington, D.C. 2013.

[2] Yang, F. Yun, M. Yang, X. Travel Time Distributions under Interrupted Flow and Its Application to Travel Time Reliability. In *Transportation Research Board 93rd Annual Meeting*. Transportation Research Board of the National Academies, Washington, D.C. 2014

[3] Park, S. Rakha, H. Guo, F. Calibration Issues for Multistate Model of Travel Time Reliability. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2188,* Transportation Research Board of the National Academies, Washington, D.C. 2010 pp. 74-84.

[4] Chalumuri, R. S., Asakura, Y. 2012. Modelling travel time distribution under various uncertainties on Hanshin expressway of Japan. European Transport Research Review. 6:85-92

[5] Colberg, K., W. Suh, J. Anderson, S. Zinner, A. Guin, M. Hunter, and R. Guensler. Lane Bias Issues in Work Zone Travel Time Measurement and Reporting. In Transportation Research Record: Journal of the Transportation Research Board, No. 2458, pp. 78–87, 2014.

[6] J. Barcelo, et al., (2010) “Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Traffic Monitoring” Transportation Research Record No. 2175, Accessed from http://trb.metapress.com/content/17x1v88n6p8363p1/fulltext.pdf

[7] D. Chakraborty, *et al.*, (2006) "Discovery and Delay Analysis of Bluetooth Devices," 7th International Conference on Mobile Data Management, Accessed from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1630650

[8] B. Peterson, *et al.*, (2006) “Bluetooth Inquiry Time Characterization And Selection”, IEEE Transactions On Mobile Computing, Vol. 5, No. 9. Accessed from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1661527

[9] C. M. Day, *et al.*, "Evaluation of Arterial Signal Coordination: Methodologies for Visualizing High-Resolution Event Data and Measuring Travel Time," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2192, pp. 37-49, December 1, 2010.

[10] T. M. Brennan, *et al.*, "Influence of Vertical Sensor Placement on Data Collection Efficiency from Bluetooth MAC Address Collection Devices," *Journal of Transportation Engineering,* vol. 136, pp. 1104-1109, December 2010.

[11] P. J. Tarnoff, *et al.*, "Continuing Evolution of Travel Time Data Information Collection and Processing," in *Transportation Research Board 88th Annual Meeting*, Washington DC, 2009, p. 14.

[12] Y. Wang, *et al.*, "Field Experiments with Bluetooth Sensors," ed, 2010.

[13] Benaglia, T. Chauveau, D. Hunter, D. Young, D. 2009. mixtools: An R Package for Analyzing Finite Mixture Models. In *Journal of Statistical Software*. October 2009. Volume 32, Issue 6.

[14] *Package 'mixtools'.* CRAN, Derek Young, April 19, 2015.

[15] K. P. Burnham *et al.*, "AIC Model Selection and multimodel inference in behavioral ecology: some background, observations, and comparisons," in *Behavioral Ecology and Sociobiology,* Vol. 65, no. 1, 2011, pp. 23-35